Electronic Journal of Probability

Electron. J. Probab. **21** (2016), no. 8, 1–42.
ISSN: 1083-6489 DOI: 10.1214/16-EJP4180

Asymptotic entropy of random walks on regular languages over a finite alphabet

Lorenz A. Gilch*

Abstract

We prove existence of asymptotic entropy of random walks on regular languages over a finite alphabet and we give formulas for it. Furthermore, we show that the entropy varies real-analytically in terms of probability measures of constant support, which describe the random walk. This setting applies, in particular, to random walks on virtually free groups.

Keywords: random walks; regular languages; entropy; analytic.
AMS MSC 2010: Primary 60J10, Secondary 28D20.
Submitted to EJP on March 19, 2015, final version accepted on January 19, 2016.
Supersedes arXiv:1304.3555.

1 Introduction

Let $\mathcal{A}$ be a finite alphabet and let $\mathcal{A}^*$ be the set of all finite words over the alphabet $\mathcal{A}$, where o denotes the empty word. Consider a transient Markov chain $(X_n)_{n\in\mathbb{N}_0}$ on $\mathcal{A}^*$ with $X_0 = o$ such that at each instant of time the last $K \in \mathbb{N}$ letters of the current word may be replaced by a word of length of at most $2K$ and the transition probabilities depend only on the last K letters of the current word and on the replacing word. For better visualization and ease of presentation, we also consider the random walk on $\mathcal{A}^*$ as a random walk on an undirected graph $\mathcal{G}$. Denote by π_n the distribution of X_n. We are interested whether the sequence $\frac{1}{n}\mathbb{E}[-\log \pi_n(X_n)]$ converges, and if so to describe the limit. If it exists, it is called the *asymptotic entropy*, which was introduced by Avez [1]. The aim of this paper is to prove existence of the asymptotic entropy, to describe it as the rate of escape w.r.t. the Greenian distance and to prove its real-analytic behaviour when varying the transition probabilities of constant support.

We outline some background on this topic. Random Walks on regular languages have been studied by e.g. Lalley [16] and Malyshev [19] amongst others. Concerning asymptotic entropy it is well-known by Kingman's subadditive ergodic theorem (see Kingman [15]) that the entropy exists for random walks on groups if $\mathbb{E}[-\log \pi_1(X_1)] < \infty$. In contrast to this fact existence of the entropy on more general structures is not known

*Graz University of Technology, Austria. E-mail: Lorenz.Gilch@freenet.de http://www.math.tugraz.at/~gilch/

a priori. In our setting we are not able to apply the subadditive ergodic theorem since we neither have subadditivity nor a global composition law of words if the random walk is performed on a proper subset of $\mathcal{A}^*$ (that is, not every word $w \in \mathcal{A}^*$ can be reached from o with positive probability). This forces us to use other techniques like generating functions techniques. These generating functions are power series with probabilities as coefficients, which describe the characteristic behaviour of the underlying random walks. The technique of our proof of existence of the entropy was motivated by Benjamini and Peres [2], where it is shown that for random walks on groups the entropy equals the rate of escape w.r.t. the Greenian distance; compare also with Blachère, Haïssinsky and Mathieu [3]. In particular, we will also show that the asymptotic entropy h is the rate of escape w.r.t. a distance function in terms of Green functions, which in turn yields that h is also the rate of escape w.r.t. the Greenian distance. Moreover, we prove convergence in probability and convergence in L_1 of the sequence $-\frac{1}{n}\log \pi_n(X_n)$ to h, and we show also that h can be computed along almost every sample path as the limes inferior of the aforementioned sequence. The question of almost sure convergence of $-\frac{1}{n}\log \pi_n(X_n)$ to some constant h, however, remains open. Similar results concerning existence and formulas for the entropy are proved in Gilch and Müller [9] for random walks on directed covers of graphs and in Gilch [8] for random walks on free products of graphs. Furthermore, we give formulas for the entropy which allow numerical computations and also exact calculations in some special cases. The main idea in our proofs is to fix a priori a sequence of nested cones in the associated graph $\mathcal{G}$ and to track the random walk's way to infinity through these cones. Similar ideas have been used independently by Woess [23] for context-free pairs of groups. The techniques in our proofs are restricted to the case of bounded range random walks: in the case of unbounded range the situation gets much more complicated since Martin and Gromov boundaries may differ even under assumption of some exponential moments to be finite; compare with Gouëzel [10].

Kaimanovich and Erschler asked whether drift and entropy of random walks vary continuously (or even analytically) when varying the probabilities of the random walk with keeping the support of single step transitions constantly. In view of this question we also show in this article that h is real-analytic in terms of the parameters describing the random walk on $\mathcal{A}^*$. This fact applies, in particular, to the case of bounded range random walks on virtually free groups, which goes beyond the scope of previous results related to the question of analyticity. Ledrappier [17] showed that the entropy varies real-analytically for finitely supported random walks on free groups; with the help of "barriers" (that is, nested sequences of subsets which have to be passed successively) and the study of Martin kernels he identifies the entropy as the boundary entropy. The present article uses also some kind of barriers (called "cones") to track the random walk's path to infinity, but the approach is different: here, we identify the entropy as the Shannon entropy of a hidden Markov chain (see Theorem 2.5), which arises from splitting up the random walk into pieces between the entries of these nested cones. For some special cases (e.g., free groups) we even give a formula (see Theorem 7.4) for the entropy of the hidden Markov chain, which allows numerical calculations. A similar idea for proving existence of the entropy has also been used in Gilch [8] for random walks on free products of graphs by cutting the random walk into pieces; Theorem 7.4 applies also to free products of *finite* graphs, but not necessarily for free products of infinite graphs. The important difference between [8] and the present article is that analyticity of the entropy in [8] follows directly from the formulas for the entropy, while we have to make much more effort to show this property in the present context of regular languages. Finally, let us remark that random walks on regular languages do not only extend results from free groups or free products to the next general case like virtually free groups but also to a wider class like context-free graphs (see Subsection 2.2).

At this point let us summarize further papers concerning continuity and analyticity of the drift and entropy that have been published recently: Ledrappier [18] showed that the drift and entropy of finitely supported random walks on hyperbolic groups are Lipschitz, while Mathieu [20] showed that the entropy of symmetric, finitely supported random walks in hyperbolic groups are differentiable; Haïssinsky, Mathieu and Müller [11] proved analyticity of the drift for random walks on surface groups. The recent survey article of Gilch and Ledrappier [6] collects several results about analyticity of drift and entropy of random walks on groups.

The basic reasoning of our proofs follows a similar argumentation as in [9] and [8], but since a straight-forward adaption is not possible we have to do more effort in the present setting: we will show that the entropy equals the rate of escape w.r.t. some special length function, and we deduce the proposed properties analogously. For the proof of analyticity of the entropy we will extract a hidden Markov chain from our random walk and we will apply a result of Han and Marcus [12]. The plan of the paper is as follows: in Sections 2 and 3 we define the random walk on $\mathcal{A}^*$ and the associated generating functions. Section 4 explains the construction of cones in the present context. In Sections 5 and 6 we prove existence of the asymptotic entropy and give a formula for it, while in Section 7 we give estimates and a more explicit formula in some special case. Section 8 shows real-analyticity of the entropy.

2 Random walks on regular languages

2.1 Definitions and main results

Let $\mathcal{A}$ be a finite alphabet and denote by $\mathcal{A}^*$ the set of all finite words over $\mathcal{A}$. We write o for the empty word and $\mathcal{A}^n$, $n \in \mathbb{N}$, for the set of all words over $\mathcal{A}$ consisting of exactly n letters. For two words $w_1, w_2 \in \mathcal{A}^*$, $w_1 w_2$ denotes the concatenated word. A *random walk on a regular language* is a Markov chain $(X_n)_{n\in\mathbb{N}_0}$ on the set $\mathcal{A}^* = \bigcup_{n\geq 1} \mathcal{A}^n \cup \{o\}$, whose transition probabilities obey the following rules:

(i) Only the last two letters of the current word may be modified.
(ii) Only one letter may be adjoined or deleted at one instant of time.
(iii) Adjunction and deletion may only be done at the end of the current word.
(iv) Probabilities of modification, adjunction or deletion depend only on the last two letters of the current word and on the substitute letters.

Compare with Lalley [16] and Gilch [7]. In other words, at each step the last two letters of the current word may be replaced by a non-empty word of length of at most 3 and the transition probabilities depend only on the last two letters of the current word and the replacing word of length of at most 3. More formally, the transition probabilities of the Markov chain $(X_n)_{n\in\mathbb{N}_0}$ can be written as follows, where $w \in \mathcal{A}^*$, $a_1, a_2, b_1, b_2, b_3 \in \mathcal{A}$:

$$\begin{aligned}
\mathbb{P}[X_{n+1} = wb_1b_2 \mid X_n = wa_1a_2] &= p(a_1a_2, b_1b_2),\\
\mathbb{P}[X_{n+1} = wb_1b_2b_3 \mid X_n = wa_1a_2] &= p(a_1a_2, b_1b_2b_3),\\
\mathbb{P}[X_{n+1} = wb_1 \mid X_n = wa_1a_2] &= p(a_1a_2, b_1),\\
\mathbb{P}[X_{n+1} = b_1 \mid X_n = a_1] &= p(a_1, b_1),\\
\mathbb{P}[X_{n+1} = b_1b_2 \mid X_n = a_1] &= p(a_1, b_1b_2),\\
\mathbb{P}[X_{n+1} = o \mid X_n = a_1] &= p(a_1, o),\\
\mathbb{P}[X_{n+1} = b_1 \mid X_n = o] &= p(o, b_1),\\
\mathbb{P}[X_{n+1} = o \mid X_n = o] &= p(o, o).
\end{aligned} \tag{2.1}$$

Not all of these probabilities need to be strictly positive. Initially, we set $X_0 := o$. If we start the random walk at $w \in \mathcal{A}^*$ instead of o, we write $\mathbb{P}_w[\,\cdot\,] := \mathbb{P}[\,\cdot \mid X_0 = w]$.

The n-step transition probabilities are denoted by $p^{(n)}(w_1, w_2) := \mathbb{P}_{w_1}[X_n = w_2]$ for any $w_1, w_2 \in \mathcal{A}^*$. The set of *accessible words* from o is given by

$$\mathcal{L} = \left\{ w \in \mathcal{A}^* \mid \exists n \in \mathbb{N} : \mathbb{P}[X_n = w \mid X_0 = o] > 0 \right\}.$$

We will also think of the random walk $(X_n)_{n \in \mathbb{N}_0}$ as a nearest neighbour random walk on an *undirected graph* $\mathcal{G}$, where the vertices are the elements of $\mathcal{L}$ and undirected edges are between two vertices if and only if one can walk from one word to the other one in a single step. For this purpose, we need the following assumption:

Assumption 2.1 (Weak symmetry)**.** For all $u, v \in \mathcal{A}^*$ we assume that $\mathbb{P}_u[X_1 = v] > 0$ implies $\mathbb{P}_v[X_1 = u] > 0$. We call this property *weak symmetry*.

In particular, Assumption 2.1 yields irreducibility of the random walk on $\mathcal{L}$. Moreover, this assumption will be necessary for the construction of a sequence of cones in the graph $\mathcal{G}$ which track the random walk's way to infinity. As the interested reader will see, weak symmetry can obviously be weakened in some way but for reason of better readability we keep this natural assumption; for a discussion on this assumption, we refer to Appendix A.2.

Since the purpose of the paper is the investigation of the asymptotic behaviour of transient random walks, we obviously need that $\mathcal{L}$ is infinite in our setting. It is an easy exercise to check that the set $\mathcal{L}$ is a *regular language* over the alphabet $\mathcal{A}$, that is, the words are accepted by a finite-state automaton. For more details on regular languages, we refer e.g. to Hopcraft and Ullman [13]. Since we make no further use of the theory of languages, we will not discuss this in more detail but we remark the recursive structure of regular languages. Let us note that bounded range random walks on *virtually free groups* constitute a special case of our setting, and our results directly apply; see e.g. Lalley [16]. Thus, our results apply directly to a large class of random walks on groups and go beyond recent results for random walks on groups.

Remark 2.2. Observe that the assumption that transition probabilities depend only on the *last two* letters of the current word and that changes of the current word involve only the last two letters may be weakened to dependence and changes of the last $K \in \mathbb{N}$ letters of the current word and replacements of the last K letters by words of length of at most $2K$. This is done by blocking words of length of at most K to new single letters; see [16, Section 3.3] for further details and comments. If we make further assumptions on our random walk in the following, we will show that it does not depend on the fact if we use the "blocked letter language" (that is, dependence on the last two letters as given by (2.1) after an application of the "recoding trick") or the general case (dependence on the last K letters as given by (B.1)), that is, no required properties are lost when switching from the K-dependent case to the "blocked letter language"; for further comments, see Appendix B. It will turn out that the K-dependent case works completely analogously as the "blocked letter language" case; however, the derived equations and formulas are much more complex, so we restrict ourselves onto the case where the random walk is defined as at the beginning of this section via (2.1). In particular, there is no additional gain in the techniques and proofs when investigating the K-dependent case. Finally, let us note that it is not sufficient to consider the case where the transition probabilities/changes of words involve only the last letter in order to be able to apply this recoding trick!

We introduce some notation. The *natural word length* of any $w \in \mathcal{A}^*$ is denoted by $|w|$. If $w \in \mathcal{A}^*$ and $k \in \mathbb{N}$ with $|w| \geq k$ then $w[k]$ denotes the k-th letter of w, and $[w]$ denotes the last two letters of w when $w \neq o$ is not a single letter.

Malyshev [19] proved that the rate of escape w.r.t. the natural word length exists for irreducible random walks on regular languages, that is, there is a non-negative constant

ℓ such that

$$\lim_{n\to\infty} \frac{|X_n|}{n} = \ell \quad \text{almost surely.}$$

Here, ℓ is called the *rate of escape*. Furthermore, by [19] follows that ℓ is strictly positive if and only if $(X_n)_{n\in\mathbb{N}_0}$ is transient. In [7] there are explicit formulas for the rate of escape w.r.t. more general length functions.

Another characteristic number of random walks is the asymptotic entropy. Denote by π_n the distribution of X_n. If there is a non-negative constant h such that the limit

$$h = \lim_{n\to\infty} -\frac{1}{n}\mathbb{E}\bigl[\log \pi_n(X_n)\bigr]$$

exists, then h is called the *asymptotic entropy*. Since we only have a partial composition law for concatenation of two words (if $\mathcal{L} \subset \mathcal{A}^*$) and since we have no subadditivity and transitivity of the random walk, we are not able to apply – as in the case of random walks on groups – Kingman's subadditive ergodic theorem in order to show existence of h. It is, however, easy to see that the entropy equals zero if the random walk is recurrent (see Corollary 7.2). Therefore, from now on we will only consider *transient* random walks $(X_n)_{n\in\mathbb{N}_0}$.

Remark 2.3. Observe that $\lim_{n\to\infty} -\frac{1}{n}\log\pi_n(X_n)$ is not necessarily deterministic: take two homogeneous trees of different degrees $d_1, d_2 \geq 3$; identify their root with one single root which becomes o and consider the simple random walk on this new inhomogeneous tree with starting point o. Obviously, this random walk can be modelled as a random walk on a regular language. Then the limit $\lim_{n\to\infty} -\frac{1}{n}\log\pi_n(X_n)$ depends on the fact in which of the two subtrees the random walks goes to infinity. Hence, the sequence $-\frac{1}{n}\log\pi_n(X_n)$ converges with probability $d_1/(d_1+d_2)$ to $\log(d_1-1)$ and with probability $d_2/(d_1+d_2)$ to $\log(d_2-1)$; this can, e.g., be calculated by the formulas given in [8].

We have to make another assumption on the transition probabilities:

Assumption 2.4 (Suffix-irreducibility)**.** We assume that the random walk on $\mathcal{L}$ is *suffix-irreducible*, that is, for all $w = w_0a_0b_0 \in \mathcal{L}$ with $w_0 \in \mathcal{A}^*, a_0b_0 \in \mathcal{A}^2$ and for all $ab \in \mathcal{A}^2$ there is $n \in \mathbb{N}$ and $w_1 \in \mathcal{A}^*$ such that

$$\mathbb{P}\Bigl[X_n = w_0w_1ab, \forall k \leq n : |X_k| \geq |w| \Bigm| X_0 = w\Bigr] > 0.$$

This assumption excludes degenerate cases and will guarantee existence of ℓ; compare with [7, End of Section 2.1]. We remark that famous previous papers about random walks on regular languages (in particular, the basic ones of [19] and [16]) require stronger assumptions than this non-degeneracy assumption. Later on it will be clear that one can relax this condition in some way without needing additional techniques or ideas for the proofs. Hence, for purpose of ease and better readability, we keep this assumption until further notice. We will give further comments on this assumption in Appendix A.1.

The main idea behind our proofs will be the construction of an a priori fixed sequence of cones (that is, special subsets of $\mathcal{L}$), from which we extract a subsequence of nested cones which gives the information how the random walk tends to infinity. This extraction will be done via a hidden Markov chain $(\mathbf{Y}_k)_{k\in\mathbb{N}}$ with an underlying positive recurrent Markov chain: the asymptotic entropy $H(\mathbf{Y})$ of the process $(\mathbf{Y}_k)_{k\in\mathbb{N}}$ is given by (5.6). The average distance between two nested cones will be denoted by λ which is given by (5.8): if $X_{\mathbf{e}_k}$ denotes the word (i.e., the vertex in $\mathcal{G}$) where the k-th nested subcone is finally entered with no further exits of this cone, then $\lambda = \mathbb{E}[|X_{\mathbf{e}_2}| - |X_{\mathbf{e}_1}|]$. Our first main result concerns existence of the asymptotic entropy, which is finally proven in Section 6:

Theorem 2.5. Consider a transient random walk $(X_n)_{n\in\mathbb{N}_0}$ on a regular language, which satisfies Assumptions 2.1 and 2.4. Then the asymptotic entropy h of $(X_n)_{n\in\mathbb{N}_0}$ exists and equals

$$h = \frac{\ell \cdot H(\mathbf{Y})}{\lambda},$$

where $H(\mathbf{Y})$ is given by (5.6) and λ by (5.8).

Recall that the random walk is described by the values in (2.1). A natural question is whether the entropy varies regularly if the parameters in (2.1) are varied slightly and if positive transition probabilities remain positive by this variation. The following result gives an answer to this question, where the proof is given in Section 8:

Theorem 2.6. For transient random walks on regular languages satisfying Assumptions 2.1 and 2.4, the entropy h varies real-analytically under all probability measures of constant support.

Moreover, we can also describe the asymptotic entropy in the following way:

Corollary 2.7. We have the following types of convergence:

1. For almost every trajectory of the random walk $(X_n)_{n\in\mathbb{N}_0}$,

$$h = \liminf_{n\to\infty} -\frac{1}{n}\log \pi_n(X_n).$$

2. Convergence in probability:

$$-\frac{1}{n}\log \pi_n(X_n) \xrightarrow{\mathbb{P}} h.$$

3. Convergence in L_1:

$$-\frac{1}{n}\log \pi_n(X_n) \xrightarrow{L_1} h.$$

The *Greenian distance* between two words $w_1, w_2 \in \mathcal{L}$ is defined as

$$d_{\mathrm{Green}}(w_1,w_2) := -\log \mathbb{P}[\exists n \in \mathbb{N}_0 : X_n = w_2 \mid X_0 = w_1].$$

Analogously to the situation for random walks on groups, we get the following result, which is finally proven at the end of Section 6:

Corollary 2.8. The entropy is the rate of escape with respect to the Greenian distance, that is,

$$h = \lim_{n\to\infty} -\frac{1}{n} d_{\mathrm{Green}}(o, X_n) \quad \text{almost surely.}$$

Further results are given in Section 7, where we show that $h > 0$ (Corollary 7.1) for non-degenerate transient random walks, give an inequality between entropy, drift and growth (Theorem 7.3) and give an exact formula in some special case (Theorem 7.4).

2.2 Examples

We give three classical examples for regular languages.

2.2.1 Stacks

In computer science theory *stacks* play an important role for modelling algorithms. In this setting letters represent different procedures and words are lists of procedures, which are called randomly. The last letter of the current word is the actual running procedure which may produce more subprocedures or will finish some open procedures, which in turn yields that the stack is getting larger or smaller randomly. Thus, this setting can be encoded by regular languages. Compare also with Lalley [16].

2.2.2 Virtually free groups

An important class of examples is given by *virtually free groups*, that is, groups which contain a free group as a subgroup of finite index. Let Γ be a virtually free group which contains the free group $\mathbb{F}_d$ with d generators as a subgroup of index $[\Gamma : \mathbb{F}_d] = k$. Let $\mathbb{F}_d$ be generated by the elements $a_1, a_1^{-1}, \dots, a_d, a_d^{-1}$, and let $h_1, \dots, h_k$ be representants of the k different left co-sets of Γ. That is, each element $x \in \Gamma$ can be written as

$$x_1 x_2 \dots x_m h_j,$$

where $m \in \mathbb{N}_0$, $j \in \{1, \dots, k\}$ and $x_1, \dots, x_m \in \{a_1, a_1^{-1}, \dots, a_d, a_d^{-1}\}$ such that $x_i^{-1} \neq x_{i+1}$ for all $i \in \{1, \dots, m-1\}$. Now it is clear that each group invariant, finitely supported random walk on Γ can be considered as a random walk on a regular language with alphabet $\mathcal{A} = \{a_1, a_1^{-1}, \dots, a_d, a_d^{-1}, h_1, \dots, h_k\}$ since multiplication from the right changes only a bounded number of letters at the end of the current word. Compare also with the detailed example of free products with amalgamation in [7, Section 3.1]

2.2.3 Context-free graphs

Another important class is given by *context-free graphs*, and in particular by certain *Schreier graphs*, which can also be considered as random walks on regular languages. This class justifies the study of random walks on regular languages in its own right and not only as an extension of free groups or free products. We sketch the concept of context-free graphs: consider a labelled, symmetric graph $\mathcal{G}$ with root $\mathbf{r}$. Consider the connected components of $\mathcal{G}$ after removing all vertices (and adjoint edges) which are at distance less or equal than some $n \in \mathbb{N}$ to $\mathbf{r}$. If there are only finitely many different isomorphism types as labelled graphs of these connected components then the graph is called *context-free*; see Muller and Schupp [22]. We give a short explanation why these graphs fit into the setting of regular languages: later the mindful reader will notice that our random walks are performed on some graph with finitely many different cone types (that is, finitely many different isomorphism classes of connected components after removal of all vertices at distance less or equal than n to $\mathbf{r}$). Since there are only finitely many different cone types one can deduce a finite-state automaton from the context-free graph, which accepts just the words which describe the different vertices of $\mathcal{G}$. As a specific example, consider a virtually free group, a finitely generated free subgroup and an associated Schreier graph: by Woess [23, Theorem 2.10], the Schreier graph satisfies all needed irreducibility requirements. For further details, we refer to Muller and Schupp [21], [22] and Ceccherini-Silberstein and Woess [4] and [23].

3 Generating functions

For $w_1, w_2 \in \mathcal{A}^*$, $z \in \mathbb{C}$, the *Green function* is defined as

$$G(w_1, w_2|z) := \sum_{n \geq 0} p^{(n)}(w_1, w_2) \cdot z^n$$

and the *last visit generating function* as

$$L(w_1, w_2|z) := \sum_{n \geq 0} \mathbb{P}\big[X_n = w_2, \forall m \in \{1, \dots, n\} : X_m \neq w_1 \big| X_0 = w_1\big] \cdot z^n.$$

By conditioning on the last visit to w_1, an important relation between these functions is given by

$$G(w_1, w_2|z) = G(w_1, w_1|z) \cdot L(w_1, w_2|z). \tag{3.1}$$

In the following we introduce further generating functions, which also have been used analogously in [7]. Define for $a,b,c,d,e\in\mathcal{A}$ and real $z>0$

$$H(ab,c|z):=\sum_{n\geq 1}\mathbb{P}\bigl[X_n=c,\forall m<n:|X_m|>1\bigl|X_0=ab\bigr]\cdot z^n$$

and

$$\begin{aligned}\bar{L}(ab,cde|z) &:= \sum_{n\geq 1}\mathbb{P}\bigl[X_n=cde,|X_{n-1}|=2,\forall m\in\{1,\dots,n\}:|X_m|\geq 2,\bigl|X_0=ab\bigr]\cdot z^n,\\ \overline{G}(ab,cd|z) &:= \sum_{n\geq 0}\mathbb{P}\bigl[X_n=cd,\forall m\in\{1,\dots,n\}:|X_m|\geq 2\bigl|X_0=ab\bigr]\cdot z^n.\end{aligned}$$

We write $\bar{L}(ab,cde):=\bar{L}(ab,cde|1)$. These generating functions can be computed in two steps: first, one solves the following system of equations which arises by case distinction on the first step:

$$\begin{aligned}H(ab,c|z) &= p(ab,c)\cdot z+\sum_{de\in\mathcal{A}^2}p(ab,de)\cdot z\cdot H(de,c|z)\\ &\quad+\sum_{def\in\mathcal{A}^3}p(ab,def)\cdot z\cdot\sum_{g\in\mathcal{A}}H(ef,g|z)\cdot H(dg,c|z);\end{aligned}\tag{3.2}$$

compare with [16] and [7]. The system (3.2) consists of equations of quadratic order, and therefore the functions $H(\cdot,\cdot|z)$ are algebraic, if the transition probabilities are algebraic. We now get the functions $\overline{G}(ab,cd|z)$ by solving the following linear system of equations which also arises by case distinction on the first step:

$$\begin{aligned}\overline{G}(ab,cd|z) &= \delta_{ab}(cd)+\sum_{c_1d_1\in\mathcal{A}^2}p(ab,c_1d_1)\cdot z\cdot\overline{G}(c_1d_1,cd|z)+\\ &\quad+\sum_{c_1d_1e_1\in\mathcal{A}^3}p(ab,c_1d_1e_1)\cdot z\cdot\sum_{f\in\mathcal{A}}H(d_1e_1,f|z)\cdot\overline{G}(c_1f,cd|z).\end{aligned}$$

Finally, we get

$$\bar{L}(ab,cde|z)=\sum_{a_1b_1\in\mathcal{A}^2}\overline{G}(ab,a_1b_1|z)\cdot z\cdot p(a_1b_1,cde).\tag{3.3}$$

Obviously, it is sufficient to consider only those functions $H(ab,\cdot|z)$, $\overline{G}(ab,\cdot|z)$ and $L(ab,\cdot|z)$ such that there exists some $w_0\in\mathcal{A}^*$ with $w_0ab\in\mathcal{L}$; the remaining functions do not play a role for our random walk. Moreover, one can compute the Green functions of the form $G(o,w|z)$, $w\in\mathcal{L}$ with $|w|\leq 3$, by solving

$$\begin{aligned}G(w_1,w_2|z) &= \delta_{w_1}(w_2)+\sum_{w_3\in\mathcal{A}^*:|w_3|\leq 3}p(w_1,w_3)\cdot z\cdot G(w_3,w_2|z)+\\ &\quad+\mathbb{1}_3(w_1)\cdot\sum_{cde\in\mathcal{A}^3}p(w_1[2]w_1[3],cde)\cdot z\cdot\sum_{f\in\mathcal{A}}H(de,f|z)\cdot G(w_1[1]cf,w_2|z),\end{aligned}$$

where $w_1,w_2\in\mathcal{A}^*$ with $|w_1|,|w_2|\leq 3$ and $\mathbb{1}_3(w_1):=1$, if $|w_1|=3$, and $\mathbb{1}_3(w_1):=0$ otherwise.

We also define for $ab\in\mathcal{A}^2$:

$$\xi(ab):=\mathbb{P}\bigl[\forall n\geq 0:|X_n|\geq 2\bigm| X_0=ab\bigr]=1-\sum_{f\in\mathcal{A}}H(ab,f|1).$$

When starting at a word $wab\in\mathcal{L}$, where $w\in\mathcal{A}^*$, $\xi(ab)$ is the probability that the process $(X_n)_{n\in\mathbb{N}_0}$ will not visit any words of length $|wab|-1$ or smaller. In this case the prefix

w will remain constant for the rest of the process. Observe that, for transient random walks, $\xi(ab) > 0$ for all $ab \in \mathcal{A}^2$ due to Assumption 2.4. We define a "length function" on $\mathcal{L}$ by

$$l(w) := -\log L(o, w|1) \quad \text{for } w \in \mathcal{L}. \tag{3.4}$$

For $n \geq 2$ and $a_1, \ldots, a_n \in \mathcal{A}$, the functions $L(o, a_1 \ldots a_n|z)$ can be rewritten as

$$\sum_{b, b_0, c_0 \in \mathcal{A}} L(o, b|z) \cdot z \cdot p(b, b_0 c_0) \sum_{\substack{b_1, \ldots, b_{n-2} \in \mathcal{A}, \\ c_1, \ldots, c_{n-2} \in \mathcal{A}}} \prod_{i=1}^{n-2} \bar{L}(b_{i-1}c_{i-1}, a_i b_i c_i|z) \cdot \overline{G}(b_{n-2}c_{n-2}, a_{n-1}a_n|z); \tag{3.5}$$

each path from o to $a_1 \ldots a_n$ is decomposed to the last times when the sets $\mathcal{A}, \mathcal{A}^2, \ldots, \mathcal{A}^{n-1}$ are visited, that is, the factor $\bar{L}(b_{i-1}c_{i-1}, a_i b_i c_i|z)$ corresponds to the parts of the paths from o to $a_1 \ldots a_n$ between the final exits of the sets $\mathcal{A}^i$ and $\mathcal{A}^{i+1}$.

4 Cones

4.1 Definitions of cones and properties

In this section we introduce the structure of cones in our setting. A *path* in $\mathcal{A}^*$ is a sequence of words $\langle w_0, w_1, \ldots, w_m \rangle$, $m \in \mathbb{N}$, in $\mathcal{A}^*$ such that $\mathbb{P}_{w_{i-1}}[X_1 = w_i] > 0$ for all $1 \leq i \leq m$. By weak symmetry, we have that, for each such path, the reversed sequence of words $\langle w_m, w_{m-1}, \ldots, w_0 \rangle$ is also a path. For $n \in \mathbb{N}$, define $\mathcal{A}^*_{\geq n} := \{w \in \mathcal{A}^* \big| |w| \geq n\}$. For any $w_0 \in \mathcal{A}^*_{\geq 2}$, we define the *cone* rooted at w_0 as

$$C(w_0) := \left\{ w \in \mathcal{A}^*_{\geq |w_0|} \;\middle|\; \begin{array}{c} \exists m \in \mathbb{N}_0 \; \exists \text{ path } \langle w_0, w_1, \ldots, w_{m-1}, w \rangle \\ \text{with } w_1, \ldots, w_{m-1} \in \mathcal{A}^*_{\geq |w_0|} \end{array} \right\}.$$

In other words, when we consider the associated graph $\mathcal{G}$ then the cone $C(w_0)$ can be viewed as the subgraph of $\mathcal{G}$ which is the connected component containing w_0 after removing all vertices $w' \in \mathcal{A} \setminus \mathcal{A}^*_{\geq |w_0|}$ and the adjacent edges to these w'. In particular, we have $w_0 \in C(w_0)$. If $w_1 \in C(w_0)$ then we have $C(w_1) \subseteq C(w_0)$: indeed, let be $w_2 \in C(w_1)$; therefore, $|w_2| \geq |w_1| \geq |w_0|$ and there are paths $\langle w_0, w'_1, \ldots, w'_k, w_1 \rangle$ through words $w'_1, \ldots, w'_k \in \mathcal{A}^*_{\geq |w_0|}$ and $\langle w_1, w''_1, \ldots, w''_l, w_2 \rangle$ through words $w''_1, \ldots, w''_l \in \mathcal{A}^*_{\geq |w_1|} \subseteq \mathcal{A}^*_{\geq |w_0|}$. Hence, there is a path $\langle w_0, w'_1, \ldots, w'_k, w_1, w''_1, \ldots, w''_l, w_2 \rangle$ through words in $\mathcal{A}^*_{\geq |w_0|}$, that is, $w_2 \in C(w_0)$ yielding $C(w_1) \subseteq C(w_0)$. The cone $C(w_1)$ is then called a *subcone* of $C(w_0)$.

Observe that each element $w \in C(w_0)$ has the form $w = a_1 \ldots a_{m-2}\bar{w}$, where $w_0 = a_1 \ldots a_m$ with $m \geq 2$, $a_1, \ldots, a_m \in \mathcal{A}$ and where $\bar{w} \in \mathcal{A}^*_{\geq 2}$: indeed, by definition each $w \in C(w_0)$ can be reached from w_0 by a path through words of length bigger or equal than $|w_0|$. Thus, the first $m-2$ letters are *not* changed along such a path.

By the suffix-irreducibility Assumption 2.4, we have the following important property for cones: let be $w \in \mathcal{A}^*$ and $ab, cd \in \mathcal{A}^2$; then the cone $C(wab)$ has a proper subcone $C(wxcd) \subset C(wab)$ with a suitable choice of $x \in \mathcal{A}^* \setminus \{o\}$.

Recall that $[w]$ denotes the last two letters of a word $w \in \mathcal{A}^*_{\geq 2}$. We say that two cones $C(w_1)$ and $C(w_2)$, $w_1, w_2 \in \mathcal{A}^*$, are *isomorphic* if $C([w_1]) = C([w_2])$. The following lemma explains why we call these cones "isomorphic". Since the proof of the following lemma is elementary, we omit the proof at this place and hand it in later in Appendix C.

Lemma 4.1. Let be $w_1 = a_1 \ldots a_m$, $w_2 = b_1 \ldots b_n \in \mathcal{A}^*_{\geq 2}$ with $a_1, \ldots, a_m, b_1, \ldots, b_n \in \mathcal{A}$ such that $C(w_1)$ and $C(w_2)$ are isomorphic. Then:

1. The mapping $\varphi : C(w_1) \to C(w_2)$ defined by

$$\varphi(a_1 \ldots a_{m-2}\bar{w}) = b_1 \ldots b_{n-2}\bar{w} \quad \text{for } \bar{w} \in \mathcal{A}^*_{\geq 2} \text{ with } a_1 \ldots a_{m-2}\bar{w} \in C(w_1)$$

is a bijection which preserves the adjacency relation, that is, $p(w',w'')>0$ if and only if $p\big(\varphi(w'),\varphi(w'')\big)>0$ for all $w',w''\in C(w_1)$.

2. The cones are isomorphic as subgraphs of $\mathcal{G}$.

The lemma says implicitly that the words of two isomorphic cones differ only by different prefixes. Moreover, there is a natural 1-to-1 correspondence of paths inside $C(w_1)$ and paths in an isomorphic cone $C(w_2)$ where obviously each such path in $C(w_1)$ and the corresponding path in the other isomorphic cone $C(w_2)$ have the same probability: let be $\langle w_0',w_1',\dots,w_m'\rangle$ a path in $C(w_1)$; then $\langle\varphi(w_0'),\varphi(w_1'),\dots,\varphi(w_m')\rangle$ is a path in $C(w_2)$ and

$$\mathbb{P}\big[X_1=w_1',\dots,X_m=w_m'\big|X_0=w_0'\big]=\mathbb{P}\big[X_1=\varphi(w_1'),\dots,X_m=\varphi(w_m')\big|X_0=\varphi(w_0')\big].$$

We remark that $C(w)$ and $C(w')$, $w,w'\in\mathcal{A}^*_{\geq 2}$, with $C([w])\neq C([w'])$ can still be isomorphic as subgraphs of $\mathcal{G}$ but we will still distinguish them as elements of different isomorphism classes according to our definition of isomorphism of cones.

Our construction of cones ensures that different cones are either nested in each other or disjoint as the next lemma will show; the elementary proof of the next lemma is again omitted and will be handed in later in the Appendix C.

Lemma 4.2. Let be $w_1,w_2\in\mathcal{A}^*_{\geq 2}$. Then the cones $C(w_1)$ and $C(w_2)$ are either nested in each other, that is, $C(w_1)\subseteq C(w_2)$ or $C(w_2)\subseteq C(w_1)$, or they are disjoint, that is, $C(w_1)\cap C(w_2)=\emptyset$. If we even have $|w_1|=|w_2|$ then we have $C(w_1)=C(w_2)$ or $C(w_1)\cap C(w_2)=\emptyset$.

At this point let us mention that the weak symmetry Assumption 2.1 is crucial here: if this assumption is dropped then two cones $C(w_1)$ and $C(w_2)$, where $w_1,w_2\in\mathcal{A}^*_{\geq 2}$ with $|w_1|=|w_2|$ and $C(w_1)\cap C(w_2)\neq\emptyset$ may be non-isomorphic. This case makes everything much more difficult in our proofs since the property of cones from the last lemma (either nestedness or disjointness) is lost and since we want to track the random walk's way to infinity by distinguishing which of the (disjoint) cones are successively finally entered on its way to infinity. The author is however confident that one can adapt the situation if weak symmetry does not hold but this would need much more effort with loss of good readability of our proofs and no additional gain of the techniques; for further comments see Appendix A.2.

Since isomorphism of cones depends only on the last two letters of their roots, we have obviously only finitely many different isomorphism classes of cones. These isomorphism classes can be described by two-lettered words $ab\in\mathcal{A}^2$: first, for each isomorphism class of cones we fix some ab representing the class of $C(ab)$. Let $\mathcal{J}\subseteq\mathcal{A}^2$ be a system of representants of the different isomorphism classes of cones. Thus, for every $w\in\mathcal{A}^*_{\geq 2}$ there is some unique $ab\in\mathcal{J}$ such that $C([w])=C(ab)$. Then we write $\tau\big(C(w)\big):=ab$ for the *cone type* (or *isomorphism class*) of the cone $C(w)$. . The *boundary* of $C(w)$ is given by the set

$$\partial C(w)=\big\{w_0\in C(w)\,\big|\,|w_0|=|w|,\exists w'\in\mathcal{A}^*\setminus C(w):p(w,w')>0\big\}.$$

We have $\{[w]\mid w\in\partial C(w_1)\}=\{[w]\mid w\in\partial C(w_2)\}$ for two ismorphic cones $C(w_1)$ and $C(w_2)$ with $w_1,w_2\in\mathcal{A}^*_{\geq 2}$, which follows from the following fact: if $x_1\in\partial C(w_1)$ and $w'\in\mathcal{A}^*\setminus C(w_1)$ with $p(x_1,w')>0$, then there is, due to 4.1.(1), some $x_2\in C(w_2)$ with $[x_1]=[x_2]$ and $p([x_2],a)=p([x_1],a)>0$, where $a\in\mathcal{A}$ is the last letter of w'. This implies existence of some $w''\in\mathcal{A}^*\setminus C(w_2)$ with $p(x_2,w'')>0$.

We say that the graph $\mathcal{G}$ is *expanding* if each cone $C(w_0)$, $w_0\in\mathcal{L}$, contains two proper disjoint subcones, that is, if there exist subcones $C(w_1),C(w_2)\subsetneq C(w_0)$, $w_1,w_2\in\mathcal{L}$, with $C(w_1)\cap C(w_2)=\emptyset$. We call the random walk *expanding* if the associated graph $\mathcal{G}$ is expanding. The results below do *not* depend on whether the random walk is expanding or not. At the end, however, we will see that the non-expanding case leads to zero entropy.

Finally, let us remark that in the case of K-dependent random walks on $\mathcal{A}^*$ suffix-irreducibility can be defined analogously and cones can be defined in the exactly same way; the different cone types would be defined by words of length K. In Appendix B we will check that suffix-irreducibility and the the "expanding" property are inherited by the blocked letter language if these properties are satisfied for the K-dependent random walk.

4.2 Covering of cones by subcones

The next task is to cover (up to a finite complement) any cone $C(w)$, $w \in \mathcal{L}$, by a finite set of pairwise disjoint subcones $C_1, \dots, C_{n(w)} \subset C(w)$ such that

$$\{\tau(C_1), \dots, \tau(C_{n(w)})\} = \mathcal{J} \quad \text{and} \quad \Big| C(w) \setminus \bigcup_{i=1}^{n(w)} C_i \Big| < \infty,$$

that is, every cone type appears among these subcones and the subcones cover $C(w)$ up to finitely many words. We then call $C_1, \dots, C_{n(w)}$ a *covering* of $C(w)$. In the next subsection we show how to construct this covering when $\mathcal{G}$ is expanding; in Subsection 4.2.2 we consider the case when $\mathcal{G}$ is *not* expanding.

4.2.1 Covering for expanding random walks

Suppose we are given a cone $C(w)$ with $w = w_0 a_0 b_0 \in \mathcal{L}$, where $w_0 \in \mathcal{A}^*$ and $a_0 b_0 \in \mathcal{A}^2$. Inside this cone we can find subcones of the form $C(w_0 w' ab)$ for *each* $ab \in \mathcal{A}^2$ with suitable $w' \in \mathcal{A}^* \setminus \{o\}$ depending on ab due to suffix-irreducibility. Now we want to find subcones of each type $ab \in \mathcal{J}$ which are even pairwise disjoint. We proceed as follows to find these pairwise disjoint cones of all types: since we assume in this subsection that $\mathcal{G}$ is expanding there are paths from $w = w_0 a_0 b_0$ inside $\mathcal{A}^*_{\geq |w|}$ to words $w_0 w_1 a_1 b_1$ and $w_0 w_2 a_2 b_2$, where $w_1, w_2 \in \mathcal{A}^* \setminus \{o\}$, $a_1 b_1, a_2 b_2 \in \mathcal{A}^2$ and $C(w_0 w_1 a_1 b_1) \cap C(w_0 w_2 a_2 b_2) = \emptyset$. Then we have found a subcone of type $\tau(C(a_1 b_1))$, and we search for other cone types in the subcone $C(w_0 w_2 a_2 b_2)$ in the same way. Obviously, a subcone in $C(w_0 w_2 a_2 b_2)$ does not intersect $C(w_0 w_1 a_1 b_1)$. Iterating this step leads to a finite set $\{C_1, \dots, C_{|\mathcal{J}|}\}$ of subcones of $C(w)$ such that $\{\tau(C_1), \dots, \tau(C_{|\mathcal{J}|})\} = \mathcal{J}$ and $C_i \cap C_j = \emptyset$ for $i, j \in \{1, \dots, |\mathcal{J}|\}$ with $i \neq j$. After we have found these non-intersecting subcones of all types in $C(w)$ we cover the cone $C(w)$ by further disjoint subcones: let be $D = 1 + \max\{|w'| \mid w' \in \bigcup_{i=1}^{|\mathcal{J}|} \partial C_i\}$; define $M_D = \{w' \in C(w) \mid |w'| = D\}$. Then we can choose a subset $M := \{w'_1, \dots, w'_k\} \subseteq M_D$ such that for all $i, j \in \{1, \dots, k\}$ with $i \neq j$ and all $n \in \{1, \dots, |\mathcal{J}|\}$ we have: $C(w'_i) \cap C_n = \emptyset$, $C(w'_i) \cap C(w'_j) = \emptyset$ and

$$C(w) \setminus \left(\bigcup_{m=1}^{|\mathcal{J}|} C_m \cup \bigcup_{n=1}^{k} C(w'_n) \right)$$

is finite. This is done as follows: write $M_D := \{x_1, \dots, x_N\}$ and set $M_0 := \emptyset$. For every $i \in \{1, \dots, N\}$, perform the following steps with increasing i: if $x_i \in \bigcup_{j=1}^{|\mathcal{J}|} C_j \cup \bigcup_{x \in M_{i-1}} C(x)$, then drop x_i and set $M_i := M_{i-1}$. Otherwise, set $M_i := M_{i-1} \cup \{x_i\}$. In the latter case we cannot have $C_j \subset C(x_i)$ for some $j \in \{1, \dots, |\mathcal{J}|\}$ due to the choice of D (words in ∂C_j have word length smaller than D and all words in $C(x_i)$ have length of at least D) and also not $C(x_i) \subset C_j$, which would lead to the contradiction $x_i \in C_j$ otherwise. We also cannot have $C(x_j) \subset C(x_i)$ for $j < i$ because this implies, by Lemma 4.2, $C(x_i) = C(x_j)$ and therefore $x_i \in C(x_j)$. At the end of this procedure we get some M_N and set $M := M_N$. Since every path from w to infinity inside $C(w)$ has to pass through a word of length D we have ensured that each $w' \in C(w)$ with $|w'| = D$ lies in one of the cones $C_1, \dots, C_{\mathcal{J}}, C(x)$, $x \in M$. Thus, the set $C(w) \setminus \bigcup_{m=1}^{|\mathcal{J}|} C_m \cup \bigcup_{x \in M} C(x)$ is finite and the covering of $C(w)$ is given by the subcones

$$C_1, \dots, C_{|\mathcal{J}|}, C(x), x \in M.$$

See Figure 1 for better visualization.

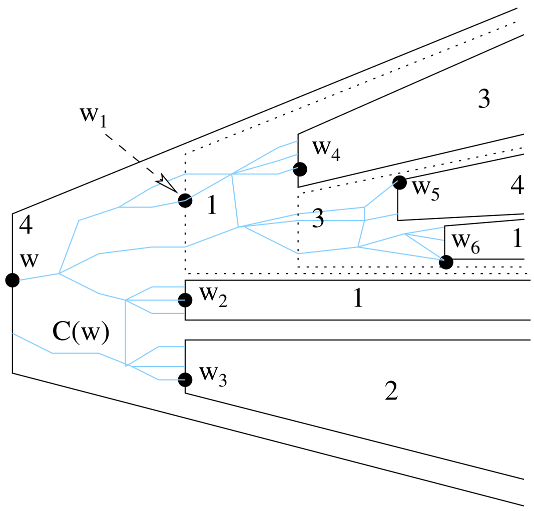

Figure 1: Covering of cones by subcones: the numbers represent the four different cone types; the cones with the solid boundary lines belong to the covering of $C(w)$. The construction of a covering is done as follows: e.g., we find three three cones in $C(w)$ whose union covers $C(w)$ up to a finite set, say the cones $C(w_1)$ (type 1), $C(w_2)$ (type 1) and $C(w_3)$ (type 2). We keep the cones $C(w_2)$ and $C(w_3)$ for the covering of $C(w)$ and search for cones of type 3 and 4 in the subcone $C(w_1)$. After having found cones of type 3 and 4 in $C(w_1)$ (for instance, the cones $C(w_4)$ and $C(w_5)$) we take additional disjoint cones in $C(w_1)$ (in the picture the innermost type-1 cone $C(w_6)$ only) into the covering such that the complement of the union of all subcones in the covering is finite. That is, the covering of $C(w)$ consists of the cones $C(w_2)$, $C(w_3)$, $C(w_4)$, $C(w_5)$ and $C(w_6)$.

The crucial point now is that we fix a covering for each cone type such that the relative positions of the subcones in the covering of some cone $C(w)$ do *not* depend on the choice of the specific root $w \in \mathcal{L}$ on the boundary of $C(w)$ but only on $\tau(C(w))$: first, for each $ab \in \mathcal{J}$, choose any $w_{ab} \in \mathcal{A}^*$ such that $w_{ab}ab \in \mathcal{L}$ and fix some covering for $C(w_{ab}ab)$, say the cones $C(w_{ab}v_1), \dots, C(w_{ab}v_k)$, where $v_1, \dots, v_k \in \mathcal{A}^*_{\geq 3}$. If $w = w_0a_1b_1 \in \mathcal{L}$ with $w_0 \in \mathcal{A}^*$, $a_1b_1 \in \mathcal{A}^2$ and $\tau(C(w)) = ab = \tau(C(w_{ab}ab))$ then we set the covering of $C(w)$ as the one which is inherited from the covering of $C(w_{ab}ab)$ by the relative location of the subcones, that is, we set the covering of $C(w)$ as the set of subcones $C(w_0v_1), \dots, C(w_0v_k)$.

Lemma 4.3. The set of subcones $C(w_0v_1), \dots, C(w_0v_k)$ is a covering of $C(w)$.

Proof. First, $C(w_0v_1), \dots, C(w_0v_k)$ are subcones of $C(w)$ since $ab \in C([w])$ (yielding $w_0ab \in \partial C(w)$) and due to the following conclusion: for each $i \in \{1, \dots, k\}$, there is a path from $w_{ab}ab$ to $w_{ab}v_i$ through words in $\mathcal{A}^*_{\geq |w_{ab}ab|}$, which implies that there is a path from ab to v_i through words in $\mathcal{A}^*_{\geq 2}$ yielding existence of a path from $w = w_0[w]$ via w_0ab to w_0v_i through words in $\mathcal{A}^*_{\geq |w|}$. That is, $C(w_0v_i) \subset C(w)$.

Since $\mathcal{J} = \{\tau(C(v_1)), \ldots, \tau(C(v_k))\}$ the set of subcones $\{C(w_0v_1), \ldots, C(w_0v_k)\}$ contains all different types. The next step is to show disjointness of the cones $C(w_0v_1)$, $\ldots, C(w_0v_k)$. Assume w.l.o.g. that $C(w_0v_1) \subsetneq C(w_0v_2)$. Then there exists a path from w_0v_2 to w_0v_1 through words in $\mathcal{A}^*_{\geq |w_0v_2|}$. This implies that there exists a path from v_2 to v_1 through words in $\mathcal{A}^*_{\geq |v_2|} \subseteq \mathcal{A}^*_{\geq 3}$, which implies that there exists a path from $w_{ab}v_2$ to $w_{ab}v_1$ through words in $\mathcal{A}^*_{\geq |w_{ab}v_2|}$ yielding $C(w_{ab}v_1) \subseteq C(w_{ab}v_2)$, a contradiction to the choice of $C(w_{ab}v_1)$, $C(w_{ab}v_2)$ in the covering of $C(w_{ab}ab)$. Thus, the cones $C(w_0v_1), \ldots, C(w_0v_k)$ are pairwise disjoint.

Analogously, we show that $C(w) \setminus \bigcup_{i=1}^k C(w_0v_i)$ is finite. Assume that this set difference is *not* finite. Then for every $N \in \mathbb{N}$ with $N \geq 3$, there exists some $\bar{w}_N \in \mathcal{A}^*$ with $|\bar{w}_N| = N$ and $w_0\bar{w}_N \in \mathcal{A}^* \cap \overline{\bigcup_{i=1}^k C(w_0v_i)}$ such that there is a path from $w = w_0[w]$ to $w_0\bar{w}_N$ through words in $\mathcal{A}^*_{\geq |w|}$. Since $[w] \in C(ab)$ there is a path from ab to $[w]$ through words in $\mathcal{A}^*_{\geq 2}$ implying that there exists a path from ab to $\bar{w}_N \in \overline{\bigcup_{i=1}^k C(v_i)}$ through words in $\mathcal{A}^*_{\geq 2}$. But this implies that there exists a path from $w_{ab}ab$ to $w_{ab}\bar{w}_N \in \overline{\bigcup_{i=1}^k C(w_{ab}v_i)}$ through words in $\mathcal{A}^*_{\geq |w_{ab}ab|}$. This gives a contradiction since $C(w_{ab}ab) \setminus \bigcup_{i=1}^k C(w_{ab}v_i)$ is finite and therefore N cannot be large. This yields the claim. $\square$

Hence, the covering of a cone depends only on its cone type, which describes the relative location of its subcones in its interior.

We can also cover $\mathcal{L}$ (up to a finite set) by a finite number of non-intersecting subcones, where each cone type appears. To this end, we just apply the algorithm explained above and take pairwise disjoint cones of the form $C(w)$ with $w \in \mathcal{L}$ and $|w| \geq 2$. We denote by $C_1^{(0)}, \ldots, C_{n_0}^{(0)}$ the covering of $\mathcal{L}$, which contains all types in $\mathcal{J}$ and which satisfies $\left|\mathcal{L} \setminus \bigcup_{i=1}^{n_0} C_i^{(0)}\right| < \infty$.

4.2.2 Non-expanding random walks

Now we explain how to proceed if $\mathcal{G}$ is *not* expanding, that is, there is a cone $C(w)$, $w \in \mathcal{L}$, which does *not* contain two proper disjoint subcones. Recall that due to suffix-irreducibility there is, for every $ab \in \mathcal{J}$, a subcone $C(w_1) \subset C(w)$ with $[w_1] = ab$. Thus, all cones do *not* have two proper disjoint subcones, because otherwise we get a contradiction to the choice of w. This non-expanding case may, in particular, occur if $\mathcal{L}$ is a proper subset of $\mathcal{A}^*$. Take now disjoint cones $C(a_1b_1), \ldots, C(a_db_d)$, where $d \in \mathbb{N}$, $a_1b_1, \ldots, a_db_d \in \mathcal{A}^2$ with $C(a_ib_i) \cap C(a_jb_j) = \emptyset$ for all $i, j \in \{1, \ldots, d\}$ with $i \neq j$ and $\mathcal{L} \setminus \bigcup_{k=1}^d C(a_kb_k)$ is finite. As already mentioned above the cones $C(a_ib_i)$, $i \in \{1, \ldots, d\}$, do *not* contain two proper disjoint subcones. Thus, we can then cover any cone $C(w)$, $w \in \mathcal{A}^*_{\geq 2}$, by the subcone $C(w_1)$ for any $w_1 \in C(w)$ with $|w_1| = |w| + 1$ and $p(w, w_1) > 0$.

Example 4.4. In order to illustrate this situation we give a short example for this case: let $\mathcal{A} = \{a, b\}$, $p(o,a) = p(a,o) = p(o,b) = p(b,o) = p(a,ab) = p(b,ba) = \frac{1}{2}$ and $p(ab,aba) = \frac{2}{3}, p(ba,b) = \frac{1}{3}$, $p(ba,bab) = \frac{3}{4}, p(ab,a) = \frac{1}{4}$. The set $\mathcal{L}$ is then given by all words of the form $ababa\ldots ba$, $ababa\ldots bab$, $baba\ldots bab$ and $baba\ldots baba$. The random walk is transient and satisfies the Assumptions 2.1 and 2.4. We have $C(ab) \cap C(ba) = \emptyset$ and $C(ab) = C(aba) \cup \{ab\}$ and $C(ba) = C(bab) \cup \{ba\}$.

The next step is to show that a non-expanding random walk converges to one of finitely many infinite words. More precisely, since we consider transient random walks, $|X_n|$ tends almost surely to infinity. Therefore, we must have that the prefixes of arbitrary length of X_n stabilize for n large enough, that is, for each $N \in \mathbb{N}$ there exists almost surely some index $n_N \in \mathbb{N}$ such that the prefixes of length N of $X_{n_N}, X_{n_N+1}, X_{n_N+2}, \ldots,$ remain constant forever. Thus, $(X_n)_{n \in \mathbb{N}_0}$ tends to some infinite (random) word $X_\infty \in \mathcal{A}^{\mathbb{N}}$.

Lemma 4.5. If $(X_n)_{n\in\mathbb{N}_0}$ is non-expanding, then the support of X_∞ is finite.

Proof. First, assume that X_∞ starts with positive probability with the letter $a_0 \in \mathcal{A}$. Assume also that $\mathbb{P}[\forall n \geq 1 : X_n \in C(a_0b_0c_0) \mid X_0 = a_0b_0c_0] > 0$ for some $b_0c_0 \in \mathcal{A}^2$ with $a_0b_0c_o \in \mathcal{L}$. We denote by A the event that X_∞ starts with the letter a_0 and that the random walk finally enters $C(a_0b_0c_0)$ on its way to infinity. Then $\mathbb{P}[A] > 0$. On this event A, assume now that the random walk tends with positive probability to some infinite words with prefixes wa_1 and wa_2, where $w \in \mathcal{A}^*_{\geq 2}$ starts with the letter a_0 and $a_1, a_2 \in \mathcal{A}$ with $a_1 \neq a_2$. Then there must be words $wa_1b_1c_1, wa_2b_2c_2 \in C(a_0b_0c_0)$, $b_1c_1, b_2c_2 \in \mathcal{A}^2$, such that

$$\mathbb{P}\big[\exists n \in \mathbb{N} : X_n = wa_ib_ic_i, \forall m \geq n : X_m \in C(wa_ib_ic_i) \,\big|\, A\big] > 0 \text{ for } i \in \{1,2\}.$$

Obviously, $C(wa_1b_1c_1) \cap C(wa_2b_2c_2) = \emptyset$. But this leads to the contradiction that $C(a_0b_0c_0)$ has two proper disjoint subcones. Therefore, $C(wa_1b_1c_1) \cap \mathcal{L} = \emptyset$ or $C(wa_2b_2c_2) \cap \mathcal{L} = \emptyset$, yielding that the letter a_1 (or a_2) is deterministic on the event A. By induction, the infinite limiting word X_∞ is deterministic on the event A, and it depends only on a_0 and b_0c_0. Since there are only finitely many possibilities for a_0 and b_0c_0, the limiting word X_∞ can only take finitely many values. □

The last lemma and suffix-irreducibility directly imply that the support of the random walk is a proper subset of $\mathcal{A}^*$ if $(X_n)_{n\in\mathbb{N}_0}$ is non-expanding. The limiting words in Example 4.4 are $ababab\ldots$ and $bababa\ldots$.

5 Last entry times

In this section we prove a law of large numbers, which turns out to describe the asymptotic entropy in the later section. For this purpose, we define last entry times (compare with [7]), for which we derive a law of large numbers. In this section we will assume that $(X_n)_{n\in\mathbb{N}_0}$ is transient and we will assume Assumptions 2.1 and 2.4, where we make explicit comments when these assumptions are essential at some points. Throughout this section, we will also use the following notations: $w_0, w_1, w_2 \in \mathcal{A}^* \setminus \{o\}$ and $a, b, c, d, a_1, b_1, a_2, b_2, \ldots \in \mathcal{A}$.

5.1 Last entry time process

We define the following *last entry times*. Let $\mathbf{e}_0$ be the first time at which the random walk visits $\bigcup_{i=1}^{n_0} \partial C_i^{(0)}$ and stays in one of the cones $C_1^{(0)}, \ldots, C_{n_0}^{(0)}$ afterwards forever, that is,

$$\mathbf{e}_0 := \inf\big\{m \in \mathbb{N}_0 \,\big|\, \exists i \in \{1, \ldots, n_0\}\, \forall n \geq m : X_n \in C_i^{(0)}\big\}.$$

In particular, $X_{\mathbf{e}_0} \in \bigcup_{i=1}^{n(0)} \partial C_i^{(0)}$ and $X_{\mathbf{e}_0 - 1} \notin \bigcup_{i=1}^{n(0)} C_i^{(0)}$. In other words, at time $\mathbf{e}_0$ the random walk finally enters one of the cones $C_i^{(0)}$ with no further exits. Inductively, if $X_{\mathbf{e}_k} = w \in \mathcal{L}$ for $k \geq 0$ and if $C(w)$ has the covering (determined only by the type of $C(w)$) consisting of the subcones $C_1^{(k)}, \ldots, C_{n(w)}^{(k)}$ as explained in Section 4, then

$$\mathbf{e}_{k+1} := \inf\big\{m > \mathbf{e}_k \,\big|\, \exists i \in \{1, \ldots, n(w)\}\, \forall n \geq m : X_n \in C_i^{(k)}\big\}.$$

In particular, $X_{\mathbf{e}_{k+1}} \in \bigcup_{i=1}^{n(w)} \partial C_i^{(k+1)}$ and $X_{\mathbf{e}_{k+1}-1} \notin \bigcup_{i=1}^{n(w)} \partial C_i^{(k)}$. Transience of $(X_n)_{n\in\mathbb{N}_0}$ yields $\mathbf{e}_k < \infty$ for all $k \in \mathbb{N}_0$ almost surely. Observe that X_n, $n \geq \mathbf{e}_k$, has the prefix w_0 if $X_{\mathbf{e}_k} = w_0ab$. Define the *relative increments* $(\mathbf{W}_k)_{k\in\mathbb{N}_0}$ between two last entry times as follows: set $\mathbf{W}_0 := X_{\mathbf{e}_0}$; for $k \geq 1$: if $X_{\mathbf{e}_{k-1}} = w_0ab$ and $X_{\mathbf{e}_k} = w_0w_1cd$, then set $\mathbf{W}_k := w_1cd$. Since we have only finitely many different cone types and the subcones of the covering of any cone C are nested at uniformly bounded distance (w.r.t. minimal path lengths) to ∂C, the random variables $\mathbf{W}_k$ can take only finitely many different values. Observe that we can reconstruct the values of the $X_{\mathbf{e}_k}$'s from the values of the $\mathbf{W}_k$'s: if $\mathbf{W}_l = w_l a_l b_l$ for $l \leq k$ then $X_{\mathbf{e}_k} = w_0 w_1 \dots w_k a_k b_k$.

For $w \in \mathcal{L}$, define

$$\mathcal{S}(w) := \bigcup_{i=1}^{n(w)} \partial C_i,$$

where $C_1, \dots, C_{n(w)}$ is the covering of $C(w)$ according to Section 4. Observe that $\mathcal{S}(w_1) = \mathcal{S}(w_2)$ if $C(w_1) = C(w_2)$. Define for $x = a_1 \dots a_k \in \mathcal{A}^*$ and $y = a_1 \dots a_{k-2} b_{k-1} b_k \dots b_{k+d} \in C(x)$ with $d \geq 1$ and $d = d(x,y) := |y| - |x|$:

$$\mathbb{L}(x,y) := \sum_{n \geq 0} \mathbb{P}\Big[X_n = y, X_{n-1} \notin C(y), \forall m \in \{1, \dots, n\} : X_m \in C(x) \Big| X_0 = x\Big].$$

If $d = 1$ then $\mathbb{L}(x,y) = \bar{L}(a_{k-1}a_k, b_{k-1}b_k b_{k+1})$. If $d \geq 2$ then $\mathbb{L}(x,y)$ can be rewritten as

$$\sum_{\substack{y_1,\dots,y_{d-1} \in \mathcal{A}^3: \\ y_i[1] = b_{k-2+i}}} \bar{L}(a_{k-1}a_k, y_1) \cdot \prod_{j=1}^{d-2} \bar{L}(y_j[2]y_j[3], y_{j+1}) \cdot \bar{L}(y_{d-1}[2]y_{d-1}[3], b_{k+d-2}b_{k+d-1}b_{k+d}); \tag{5.1}$$

the last equation follows from the fact that $\mathbb{L}(x,y)$ depends on x only by its last two letters $a_{k-1}a_k$ and by decomposition of the paths from x to y w.r.t the last times when the sets $\mathcal{A}^k$, $\mathcal{A}^{k+1}, \dots, \mathcal{A}^{k+d-1}$ are visited on the way from x to y. That is, the l-th factor in (5.1) corresponds to the part of the path from x to y between the last entry of $\mathcal{A}^*_{\geq k+l-1}$ at the word $a_1 \dots a_{k-2} b_{k-1} \dots b_{k+l-3} y_{l-1}[2] y_{l-1}[3]$ and the last entry to $\mathcal{A}^*_{\geq k+l}$ at the word $a_1 \dots a_{k-2} b_{k-1} \dots b_{k+l-2} y_l[2] y_l[3]$ (with $y_0[2]y_0[3] = a_{k-1}a_k$ and $y_d = b_{k-2}b_{k-1}b_k$). Moreover, $\mathbb{L}(x,y) = \mathbb{L}(a_{k-1}a_k, b_{k-1}b_k \dots b_{k+d})$.

If $x_1 \in \mathcal{L}$, $x_2 \in \mathcal{S}(x_1)$ and $x_3 \in \mathcal{S}(x_2)$ then

$$\mathbb{L}(x_1, x_3) = \sum_{y \in \partial C(x_2)} \mathbb{L}(x_1, y) \cdot \mathbb{L}(y, x_3)$$

by decomposition w.r.t. the last visit of the set $\partial C(x_2)$ since $C(x_3) \subset C(x_2) \subset C(x_1)$. In particular, if $\mathbb{P}[X_{\mathbf{e}_k} = x_1, X_{\mathbf{e}_{k+1}} = x_2, \dots, X_{\mathbf{e}_{k+l}} = x_{l+1}] > 0$ for $x_1, \dots, x_{l+1} \in \mathcal{L}$ then we have

$$\begin{aligned} &\mathbb{P}[X_{\mathbf{e}_k} = x_1, X_{\mathbf{e}_{k+1}} = x_2, \dots, X_{\mathbf{e}_{k+l}} = x_{l+1}] \\ = &\sum_{x_0 \in \mathcal{L} \setminus C(x_1)} G(o, x_0|1) \cdot p(x_0, x_1) \cdot \mathbb{L}(x_1, x_2) \cdot \ldots \cdot \mathbb{L}(x_l, x_{l+1}) \cdot \xi([x_{l+1}]) \end{aligned} \tag{5.2}$$

by decomposition on the final entries of the cones $C(x_1), \dots, C(x_{l+1})$. We obtain the following important observation:

Proposition 5.1. The process $\big(\mathbf{W}_k\big)_{k \geq 1}$ is a Markov chain with transition probabilities

$$q(x,y) := \begin{cases} \frac{\xi([y])}{\xi([x])} \mathbb{L}(x,y), & \text{if } y \in \mathcal{S}(x), \\ 0, & \text{otherwise.} \end{cases}$$

Proof. Let be $w_0, \dots, w_{k+1} \in \mathcal{A}^* \setminus \{o\}$ such that $w_0 \in \bigcup_{j=1}^{n_0} \partial C_j^{(0)}$, $w_{i+1} \in S(w_i)$ for all $i \in \{0, \dots, k\}$ and $\mathbb{P}[\mathbf{W}_0 = w_0, \dots, \mathbf{W}_{k+1} = w_{k+1}] > 0$. For any such sequence $\underline{w} = (w_0, \dots, w_{k+1})$, we set $x_0(\underline{w}) := w_0$ and inductively: if $x_{k-1}(\underline{w}) = y_{k-1}a_{k-1}b_{k-1}$ with

$y_{k-1} \in \mathcal{A}^*$ and $a_{k-1}b_{k-1} \in \mathcal{A}^2$ then set $x_k(\underline{w}) := y_{k-1}w_k$. That is, if $\mathbf{W}_k = w_k$ then $X_{\mathbf{e}_k} = x_k(\underline{w})$. Then:

$$
\begin{aligned}
&\mathbb{P}\big[\mathbf{W}_1 = w_1, \dots, \mathbf{W}_k = w_k\big] = \sum_{w_0 \in \bigcup_{j=1}^{n_0} \partial C_j^{(0)}} \mathbb{P}\big[\mathbf{W}_0 = w_0, \dots, \mathbf{W}_k = w_k\big] \\
= &\sum_{w_0 \in \bigcup_{j=1}^{n_0} \partial C_j^{(0)}} \mathbb{P}\big[X_{\mathbf{e}_0} = w_0, X_{\mathbf{e}_1} = x_1(\underline{w}), \dots, X_{\mathbf{e}_k} = x_k(\underline{w})\big] \\
= &\sum_{w_0 \in \bigcup_{j=1}^{n_0} \partial C_j^{(0)}} \sum_{w' \in \mathcal{L} \setminus C(w_0)} G(o, w'|1) \cdot p(w', w_0) \cdot \prod_{i=1}^{k} \mathbb{L}(x_{i-1}(\underline{w}), x_i(\underline{w})) \cdot \xi([x_k(\underline{w})]) \\
= &\sum_{w_0 \in \bigcup_{j=1}^{n_0} \partial C_j^{(0)}} \sum_{w' \in \mathcal{L} \setminus C(w_0)} G(o, w'|1) \cdot p(w', w_0) \cdot \prod_{i=1}^{k} \mathbb{L}(w_{i-1}, w_i) \cdot \xi([w_k]).
\end{aligned}
$$

The last equation arises from (5.2) by decomposing the paths by the last entries to the sets ∂C_i, where C_i denotes the cone with $X_{\mathbf{e}_i} \in \partial C_i$. Now we obtain:

$$
\begin{aligned}
&\mathbb{P}\big[\mathbf{W}_{k+1} = w_{k+1} \mid \mathbf{W}_1 = w_1, \dots, \mathbf{W}_k = w_k\big] \\
= &\frac{\mathbb{P}\big[\mathbf{W}_1 = w_1, \dots, \mathbf{W}_k = w_k, \mathbf{W}_{k+1} = w_{k+1}\big]}{\mathbb{P}\big[\mathbf{W}_1 = w_1, \dots, \mathbf{W}_k = w_k\big]} \\
= &\frac{\sum_{w_0 \in \bigcup_{j=1}^{n_0} \partial C_j^{(0)}} \sum_{w' \in \mathcal{L} \setminus C(w_0)} G(o, w'|1) \cdot p(w', w_0) \cdot \prod_{i=1}^{k+1} \mathbb{L}(w_{i-1}, w_i) \cdot \xi([w_{k+1}])}{\sum_{w_0 \in \bigcup_{j=1}^{n_0} \partial C_j^{(0)}} \sum_{w' \in \mathcal{L} \setminus C(w_0)} G(o, w'|1) \cdot p(w', w_0) \cdot \prod_{i=1}^{k} \mathbb{L}(w_{i-1}, w_i) \cdot \xi([w_k])} \\
= &\, q(x, y).
\end{aligned}
$$

□

Define the set

$$
\mathcal{W}_0 := \big\{ w \in \mathcal{A}^* \big| \exists w_0 \in \mathcal{A}^*, ab \in \mathcal{A}^2 \text{ with } \mathbb{P}[\mathbf{W}_0 = w_0 ab, \mathbf{W}_1 = w] > 0 \big\} \subseteq \mathcal{A}^*_{\geq 3}.
$$

The next lemma desribes the support of the random variables $\mathbf{W}_k$; since the proof contains only elementary, tedious calculations, we omit it at this place and hand it in later in Appendix C.

Lemma 5.2. For all $k \geq 1$, $\mathrm{supp}(\mathbb{P}[\mathbf{W}_k = \cdot]) = \mathcal{W}_0$.

With the last lemma we can show:

Lemma 5.3. The Markov chain $(\mathbf{W}_k)_{k \in \mathbb{N}}$ is positive recurrent and aperiodic.

Proof. Since $\mathcal{W}_0$ is finite it suffices to show that the process $(\mathbf{W}_k)_{k \in \mathbb{N}}$ is irreducible and aperiodic. First we show irreducibility. Let be $w_1 = w' a_1 b_1, w_2 \in \mathcal{W}_0$. Then there is some $w_0 a_0 b_0 \in \bigcup_{j=1}^{n_0} \partial C_j^{(0)}$ such that

$$
\begin{aligned}
\mathbb{P}[\mathbf{W}_1 = w_2] &\geq \mathbb{P}[X_{\mathbf{e}_0} = w_0 a_0 b_0, \mathbf{W}_1 = w_2] \\
&= \sum_{w' \in \mathcal{L} \setminus C(w_0 a_0 b_0)} G(o, w') p(w', w_0 a_0 b_0) \mathbb{L}(w_0 a_0 b_0, w_0 w_2) \xi([w_2]) > 0.
\end{aligned}
$$

In particular, $\mathbb{L}(a_0 b_0, w_2) = \mathbb{L}(w_0 a_0 b_0, w_0 w_2) > 0$. By construction of coverings, $C(a_1 b_1)$ has a subcone of type $\tau(C(a_0 b_0))$ in its covering, say the cone $C(\tilde{w})$ with $\tilde{w} \in C(a_1 b_1) \cap \mathcal{W}_0$ and $\mathbb{L}(a_1 b_1, \tilde{w}) > 0$. Then:

$$
\begin{aligned}
\mathbb{P}[\mathbf{W}_3 = w_2 \mid \mathbf{W}_1 = w_1] &\geq q(w_1, \tilde{w}) \cdot q(\tilde{w}, w_2) \qquad (5.3) \\
&= \mathbb{L}(a_1 b_1, \tilde{w}) \mathbb{L}([\tilde{w}], w_2) \frac{\xi([w_2])}{\xi(a_1 b_1)} > 0,
\end{aligned}
$$

which follows from the fact that $\mathbb{L}([\tilde{w}], w_2) > 0$ due to $[\tilde{w}] \in C(a_0b_0)$ and $\mathbb{L}(a_0b_0, w_2) > 0$ (recall the remark before Lemma 5.2). This proves irreducibility and thus positive recurrence of $(\mathbf{W}_k)_{k\in\mathbb{N}}$.

In order to see aperiodicity of the process $(\mathbf{W}_k)_{k\in\mathbb{N}}$ choose in the proof above $w_1 = w_2$, which yields that the period of $(\mathbf{W}_k)_{k\in\mathbb{N}}$ is either 1 or 2. Now let be $w \in \mathcal{W}_0$ and take any $\hat{w} \in \mathcal{W}_0$ with $q(w, \hat{w}) > 0$. Then according to (5.3) we get

$$\mathbb{P}[\mathbf{W}_4 = w, \mathbf{W}_2 = \hat{w} \mid \mathbf{W}_1 = w] = q(w, \hat{w}) \cdot \mathbb{P}[\mathbf{W}_3 = w \mid \mathbf{W}_1 = \hat{w}] > 0,$$

which implies aperiodicity. □

For sake of better identification of the cones, we now switch to a more suitable representation of cones and coverings. We identify the different cone types by numbers $\mathcal{I} := \{1, \ldots, r\} \subset \mathbb{N}$. If $C(w)$ is a cone of type $i \in \mathcal{I}$, then the covering of $C(w)$ (according to Subsection 4.2) has $n(i,j)$ subcones of type $j \in \mathcal{I}$. We denote these subcones of type j by $C_{j_{i,k}} = C_{j_{i,k}}(w) \subset C(w)$ with $1 \leq k \leq n(i,j)$ or we just identify them by $j_{i,1}, \ldots, j_{i,n(i,j)}$, which correspond to the subcones of type j with different locations inside $C(w)$. In particular, we choose this enumeration of the subcones of type j in a consistent way: if $C(w_{ab}v_m)$ belongs to the covering of $C(ab)$, $i = \tau(C(ab))$, with $C(w_{ab}v_m)$ being the k-th cone of type j in the covering of $C(ab)$ (identified by $j_{i,k}$ w.r.t. ab), then the k-th subcone of type j in the covering of any cone $C(w_0ab)$ is the subcone $C(w_0v_m)$; compare with the construction of the covering of any cone $C(w)$ starting from the covering of the cone $C(w_{ab}ab)$ in Subsection 4.2. That is, by this enumeration of subcones we ensure that the relative position of $C_{j_{i,k}}(w)$ in the interior of $C(w)$ is always the same for any $w \in \mathcal{L}$ with $i = \tau(C(w))$. We will sometimes omit the root w in the notation of the subcones when it will be clear from the context and when only the relative position of a subcone in some given cone will be of importance.

We now track the random walk's way to infinity by looking which of the cones are finally entered successively. For this purpose, define $\mathbf{i}_k := j_{i,l}$ if $\tau(C(X_{\mathbf{e}_{k-1}})) = i$ and $X_{\mathbf{e}_k} \in \partial C_{j_{i,l}}(X_{\mathbf{e}_{k-1}})$. If we set additionally $\mathbf{i}_0 := C(X_{\mathbf{e}_0})$, then the sequence $(\mathbf{i}_k)_{k\in\mathbb{N}_0}$ tracks the random walk's way to infinity.

At this point we recall the relation between $\mathbf{W}_k$ and $X_{\mathbf{e}_k}$: if $X_{\mathbf{e}_0} = \mathbf{W}_0 = w_0a_0b_0$ and $\mathbf{W}_1 = w_1a_1b_1$ then $X_{\mathbf{e}_1} = w_0w_1a_1b_1$; in general, if $X_{\mathbf{e}_{k-1}} = wa_{k-1}b_{k-1}$ and $\mathbf{W}_k = w_ka_kb_k$ then $X_{\mathbf{e}_k} = ww_ka_kb_k$. That is, there is a natural bijection of trajectories of $(\mathbf{W}_k)_{k\in\mathbb{N}_0}$ and $(X_{\mathbf{e}_k})_{k\in\mathbb{N}_0}$. In particular, the values of the $\mathbf{W}_k$'s determine the values of the $\mathbf{i}_k$'s uniquely, since the last two letters of $\mathbf{W}_{k-1}$ describe $\tau(C(X_{\mathbf{e}_{k-1}}))$ and $\mathbf{W}_k$ describes $\tau(C(X_{\mathbf{e}_k}))$ and the corresponding number in the enumeration of subcones. For a better visualization of the values of $\mathbf{i}_k$, see Figure 2.

In other words, the random variables $\mathbf{i}_k$ collect the information of the different cones which are entered successively by the random walk $(X_n)_{n\in\mathbb{N}_0}$ on its way to infinity, while the $\mathbf{W}_k$'s keep, in addition, the information where the single subcones are finally entered.

Define

$$\mathcal{W} := \left\{ (j_{m,n}, x) \;\middle|\; \begin{array}{c} x \in \mathcal{W}_0, \exists w_0 \in \mathcal{L} : \mathbb{P}[\mathbf{W}_0 = w_0, \mathbf{W}_1 = x] > 0, \\ \tau(C([w_0])) = m, \tau(C([x])) = j, 1 \leq n \leq n(m,j) \\ \text{with } x \in \partial C_{j_{m,n}}([w_0]) \end{array} \right\}.$$

In other words, $(j_{m,n}, x) \in \mathcal{W}$ if $x \in \mathcal{W}_0$ with $\tau(C(x)) = j$ and if there is $w_0a_0b_0 \in \mathcal{L}$ such that $\tau(C(a_0b_0)) = m$, $\mathbb{P}[X_{\mathbf{e}_0} = w_0a_0b_0, X_{\mathbf{e}_1} = w_0x] > 0$ and $C(x)$ being the n-th subcone of type j in the covering of $C(a_0b_0)$.

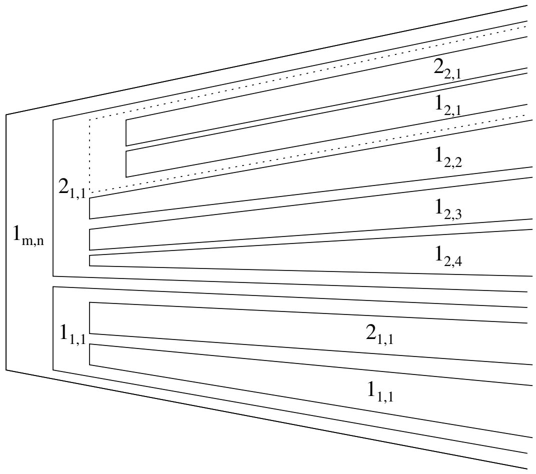

Figure 2: Numbering of subcones: the cones with the solid boundary belong to the covering while the cone with the dotted line does not.

Proposition 5.4. The process $\big((\mathbf{i}_k,\mathbf{W}_k)\big)_{k\in\mathbb{N}}$ is a positive recurrent, aperiodic Markov chain on the state space $\mathcal{W}$. Moreover, for $(i_{m,n},w_1),(j_{s,t},w_2)\in\mathcal{W}$, the transition probabilities are given by

$$\mathbb{P}\Big[(\mathbf{i}_k,\mathbf{W}_k)=(j_{s,t},w_2)\Big|(\mathbf{i}_{k-1},\mathbf{W}_{k-1})=(i_{m,n},w_1)\Big]=\begin{cases}q(w_1,w_2), & \text{if } s=i,\\ 0, & \text{if } s\neq i.\end{cases} \tag{5.4}$$

Proof. Since the values of the $\mathbf{i}_k$'s are uniquely determined by the values of the $\mathbf{W}_k$'s and since the process $(\mathbf{W}_k)_{k\in\mathbb{N}}$ is a Markov chain, we also have that $\big((\mathbf{i}_k,\mathbf{W}_k)\big)_{k\in\mathbb{N}}$ is Markovian with the proposed transition probabilities.

It remains to prove that $\mathrm{supp}(\mathbb{P}[(\mathbf{i}_k,\mathbf{W}_k)=\cdot])=\mathcal{W}$ for $k\geq 1$ and that $\big((\mathbf{i}_k,\mathbf{W}_k)\big)_{k\in\mathbb{N}}$ is positive recurrent and aperiodic. Since both proofs consist of tedious calculations analogously to the proofs of Lemmas 5.2 and 5.3 we omit these proofs here and refer to Appendix C, where we will hand in them later. $\square$

Let us recall that the values of the $\mathbf{i}_k$'s are uniquely determined by the values of the $\mathbf{W}_k$'s; however, we will explicitely keep the values of the $\mathbf{i}_k$'s in the notation of the process for sake of convenience. Observe that the process $(\mathbf{i}_k)_{k\in\mathbb{N}}$ is, in general, not Markovian. This relies on the fact that $(\mathbf{i}_k)_{k\in\mathbb{N}}$ can be seen as a function of the process $(\mathbf{W}_k)_{k\in\mathbb{N}}$: the values of the $\mathbf{W}_k$'s determine the values of the $\mathbf{i}_k$'s but not vice versa.

Define the following projection for $(i_{k,l},w_1),(j_{m,n},w_2)\in\mathcal{W}$:

$$\pi\big((i_{k,l},w_1),(j_{m,n},w_2)\big):=\begin{cases}(i,j_{i,n})=:(i,j_n), & \text{if } m=i,\\ (i,j_{i,1})=(i,j_1), & \text{if } m\neq i.\end{cases} \tag{5.5}$$

Here, j_l represents the l-th subcone of type j in the covering of a cone of type i, namely the cone represented by $j_{i,l}$. We now define the *hidden Markov chain* $(\mathbf{Y}_k)_{k\in\mathbb{N}}$ by

$$\mathbf{Y}_k:=\pi\big((\mathbf{i}_k,\mathbf{W}_k),(\mathbf{i}_{k+1},\mathbf{W}_{k+1})\big).$$

In other words, $(\mathbf{Y}_k)_{k\in\mathbb{N}}$ traces once again the random walk's way to infinity in terms of which subcones are entered successively *without* distinguishing which of the cone boundary points are the last entry time points $X_{\mathbf{e}_k}$. At this point let us mention that the second branch in the definition of $\pi(\cdot,\cdot)$ is not used for defining $\mathbf{Y}_k$, but it will be of interest in Section 8. Furthermore, observe that $X_{\mathbf{e}_0}$ and $(\mathbf{Y}_k)_{k\in\mathbb{N}}$ allow to reconstruct $(\mathbf{i}_k)_{k\in\mathbb{N}}$.

Define
$$\mathcal{W}_\pi := \bigl\{(s,t_n)\big| s,t\in\mathcal{I}, 1\leq n\leq n(s,t)\bigr\}.$$
That is, t_n corresponds to the n-th subcone of type t in the covering of a cone of type s.

Lemma 5.5. For all $k\geq 1$, $\operatorname{supp}(\mathbb{P}[\mathbf{Y}_k=\cdot])=\mathcal{W}_\pi$.

Proof. The inclusion $\operatorname{supp}(\mathbb{P}[\mathbf{Y}_k=\cdot])\subset\mathcal{W}_\pi$ is obvious by definition of $\mathbf{Y}_k$ and $\mathcal{W}_\pi$. Now we show the other inclusion. Let be $(s,t_n)\in\mathcal{W}_\pi$. Take any $w_{k-1}a_{k-1}b_{k-1}\in\mathcal{W}_0$ with $\mathbb{P}[\mathbf{W}_{k-1}=w_{k-1}a_{k-1}b_{k-1}]>0$. Then there exists $w_ka_kb_k\in\mathcal{W}_0$ with $\tau(C(a_kb_k))=s$ and $q(w_{k-1}a_{k-1}b_{k-1},w_ka_kb_k)>0$ due to the construction of coverings. Moreover, there is $w_{k+1}a_{k+1}b_{k+1}\in\mathcal{W}_0$ with $q(w_ka_kb_k,w_{k+1}a_{k+1}b_{k+1})>0$ such that $C(w_{k+1}a_{k+1}b_{k+1})$ is the n-th cone of type t in $C(a_kb_k)$. Thus,
$$\begin{aligned}
&\mathbb{P}[\mathbf{Y}_k=(s,t_n)]\\
\geq\ &\mathbb{P}[\mathbf{W}_{k-1}=w_{k-1}a_{k-1}b_{k-1},\mathbf{W}_k=w_ka_kb_k,\mathbf{W}_{k+1}=w_{k+1}a_{k+1}b_{k+1}]\\
=\ &\mathbb{P}[\mathbf{W}_{k-1}=w_{k-1}a_{k-1}b_{k-1}]\cdot q(w_{k-1}a_{k-1}b_{k-1},w_ka_kb_k)\cdot q(w_ka_kb_k,w_{k+1}a_{k+1}b_{k+1})>0,
\end{aligned}$$
yielding $(s,t_n)\in\operatorname{supp}(\mathbb{P}[\mathbf{Y}_k=\cdot])$. ☐

Since the process $(\mathbf{i}_k,\mathbf{W}_k)_{k\in\mathbb{N}}$ is positive recurrent, it has an invariant probability measure ν. Let $(\mathbf{i}_k^{(\nu)},\mathbf{W}_k^{(\nu)})_{k\in\mathbb{N}}$ be a Markov chain with transition probabilities given by (5.4) but with initial distribution ν. The corresponding hidden Markov chain $(\mathbf{Y}_k^{(\nu)})_{k\in\mathbb{N}}$ is given by
$$\mathbf{Y}_k^{(\nu)} := \pi\bigl((\mathbf{i}_k^{(\nu)},\mathbf{W}_k^{(\nu)}),(\mathbf{i}_{k+1}^{(\nu)},\mathbf{W}_{k+1}^{(\nu)})\bigr).$$
In the next section we will link the hidden Markov chains $(\mathbf{Y}_k)_{k\in\mathbb{N}}$ and $(\mathbf{Y}_k^{(\nu)})_{k\in\mathbb{N}}$.

5.2 Entropy of the hidden Markov chain related to the last entry time process

In this subsection we derive existence of the asymptotic entropy of the hidden Markov chains $(\mathbf{Y}_k^{(\nu)})_{k\in\mathbb{N}}$ and $(\mathbf{Y}_k)_{k\in\mathbb{N}}$.

First, consider the hidden Markov chain $(\mathbf{Y}_k^{(\nu)})_{k\in\mathbb{N}}$: this process is stationary and ergodic since the underlying Markov chain $\bigl(\mathbf{i}_k^{(\nu)},\mathbf{W}_k^{(\nu)}\bigr)_{k\in\mathbb{N}}$ is stationary, positive recurrent and aperiodic. Hence, there is a constant $H(\mathbf{Y})\geq 0$ such that
$$\lim_{k\to\infty}-\frac{1}{k}\log\mathbb{P}[\mathbf{Y}_1^{(\nu)}=\underline{y}_1,\ldots,\mathbf{Y}_k^{(\nu)}=\underline{y}_k]=H(\mathbf{Y}) \tag{5.6}$$
for almost every realisation $(\underline{y}_1,\underline{y}_2,\ldots)\in\mathcal{W}_\pi^{\mathbb{N}}$ of $(\mathbf{Y}_k^{(\nu)})_{k\in\mathbb{N}}$; see e.g. Cover and Thomas [5, Theorem 16.8.1]. The number $H(\mathbf{Y})$ is called the *asymptotic entropy of the (positive recurrent) process* $(\mathbf{Y}_k^{(\nu)})_{k\in\mathbb{N}}$. We now deduce an analogous statement for the process $(\mathbf{Y}_k)_{k\in\mathbb{N}}$.

Proposition 5.6. For almost every realisation $(\underline{y}_1,\underline{y}_2,\ldots)\in\mathcal{W}_\pi^{\mathbb{N}}$ of $(\mathbf{Y}_k)_{k\in\mathbb{N}}$,
$$\lim_{k\to\infty}-\frac{1}{k}\log\mathbb{P}\bigl[\mathbf{Y}_1=\underline{y}_1,\ldots,\mathbf{Y}_k=\underline{y}_k\bigr]=H(\mathbf{Y}).$$

Proof. The processes $(\mathbf{Y}_k^{(\nu)})_{k\in\mathbb{N}}$ and $(\mathbf{Y}_k)_{k\in\mathbb{N}}$ differ only by the inital distributions of $(\mathbf{i}_1^{(\nu)},\mathbf{W}_1^{(\nu)})$ and $(\mathbf{i}_1,\mathbf{W}_1)$. Moreover, there are constants $c,C>0$ such that
$$c\cdot\mathbb{P}[(\mathbf{i}_1,\mathbf{W}_1)=(i_{m,n},x)]\leq\nu(i_{m,n},x)\leq C\cdot\mathbb{P}[(\mathbf{i}_1,\mathbf{W}_1)=(i_{m,n},x)]$$
for all $(i_{m,n},x)\in\mathcal{W}$. Denote by μ_1 the distribution of $(\mathbf{i}_1,\mathbf{W}_1)$. We now get for almost every trajectory $(\underline{y}_1,\underline{y}_2,\ldots)\in\mathcal{W}_\pi^{\mathbb{N}}$ of $(\mathbf{Y}_k)_{k\in\mathbb{N}}$:

$$\begin{aligned}
& H(\mathbf{Y}) = \lim_{k\to\infty} -\frac{1}{k}\log \mathbb{P}\big[\mathbf{Y}_1^{(\nu)} = \underline{y}_1,\dots,\mathbf{Y}_k^{(\nu)} = \underline{y}_k\big] \\
= & \lim_{k\to\infty} -\frac{1}{k}\log \sum_{\substack{\underline{w}_1,\dots,\underline{w}_{k+1}\in\mathcal{W}:\\ \pi(\underline{w}_j,\underline{w}_{j+1})=\underline{y}_j\\ \text{for } 1\leq j\leq k}} \nu(\underline{w}_1)\mathbb{P}[(\mathbf{i}_l,\mathbf{W}_l)=\underline{w}_l \text{ for } 2\leq l\leq k+1 \mid (\mathbf{i}_1,\mathbf{W}_1)=\underline{w}_1] \\
= & \lim_{k\to\infty} -\frac{1}{k}\log \sum_{\substack{\underline{w}_1,\dots,\underline{w}_{k+1}\in\mathcal{W}:\\ \pi(\underline{w}_j,\underline{w}_{j+1})=\underline{y}_j\\ \text{for } 1\leq j\leq k}} \mu_1(\underline{w}_1)\mathbb{P}[(\mathbf{i}_l,\mathbf{W}_l)=\underline{w}_l \text{ for } 2\leq l\leq k+1 \mid (\mathbf{i}_1,\mathbf{W}_1)=\underline{w}_1] \\
= & \lim_{k\to\infty} -\frac{1}{k}\log \sum_{\substack{\underline{w}_1,\dots,\underline{w}_{k+1}\in\mathcal{W}:\\ \pi(\underline{w}_j,\underline{w}_{j+1})=\underline{y}_j\\ \text{for } 1\leq j\leq k}} \mathbb{P}\big[(\mathbf{i}_1,\mathbf{W}_1)=\underline{w}_1,\dots,(\mathbf{i}_{k+1},\mathbf{W}_{k+1})=\underline{w}_{k+1}\big] \\
= & \lim_{k\to\infty} -\frac{1}{k}\log \mathbb{P}\big[\mathbf{Y}_1 = \underline{y}_1,\dots,\mathbf{Y}_k = \underline{y}_k\big]. \quad \square
\end{aligned}$$

As a consequence we obtain the next statement:

Corollary 5.7.

$$\lim_{k\to\infty} -\frac{1}{k}\int \log \mathbb{P}\big[\mathbf{Y}_1 = \underline{y}_1,\dots,\mathbf{Y}_k = \underline{y}_k\big]\, d\mathbb{P}(\underline{y}_1,\underline{y}_2,\dots) = H(\mathbf{Y}).$$

Proof. Since $|\mathcal{W}|<\infty$ by definition, there is $\varepsilon_0>0$ such that, for all $\underline{w}_1,\underline{w}_2\in\mathcal{W}$,

$$\mathbb{P}[(\mathbf{i}_2,\mathbf{W}_2)=\underline{w}_2 \mid (\mathbf{i}_1,\mathbf{W}_1)=\underline{w}_1]>0 \;\text{ implies }\; 1\geq \mathbb{P}[(\mathbf{i}_2,\mathbf{W}_2)=\underline{w}_2 \mid (\mathbf{i}_1,\mathbf{W}_1)=\underline{w}_1]\geq \varepsilon_0.$$

If $(\underline{y}_1,\dots,\underline{y}_k)\in\mathcal{W}_\pi^k$ with $\mathbb{P}[\mathbf{Y}_1=\underline{y}_1,\dots,\mathbf{Y}_k=\underline{y}_k]>0$ then there are $\underline{w}_1,\dots,\underline{w}_{k+1}\in\mathcal{W}$ with $\pi(\underline{w}_j,\underline{w}_{j+1})=\underline{y}_j$ for $1\leq j\leq k$ and $\mathbb{P}\big[(\mathbf{i}_1,\mathbf{W}_1)=\underline{w}_1,\dots,(\mathbf{i}_{k+1},\mathbf{W}_{k+1})=\underline{w}_{k+1}\big]>0$. Therefore,

$$\begin{aligned}
0 &\leq -\frac{1}{k}\log \mathbb{P}\big[\mathbf{Y}_1=\underline{y}_1,\dots,\mathbf{Y}_k=\underline{y}_k\big] \\
&\leq -\frac{1}{k}\log \mathbb{P}\big[(\mathbf{i}_1,\mathbf{W}_1)=\underline{w}_1,\dots,(\mathbf{i}_{k+1},\mathbf{W}_{k+1})=\underline{w}_{k+1}\big] \\
&\leq -\frac{1}{k}\log(c\cdot\varepsilon_0^k) = -\frac{1}{k}\log c - \log\varepsilon_0 \leq -\log c - \log\varepsilon_0,
\end{aligned}$$

where $c=\min_{\underline{w}\in\mathcal{W}}\mathbb{P}[(\mathbf{i}_1,\mathbf{W}_1)=\underline{w}]$. Therefore, we may exchange integral and limit, which yields the claim together with Proposition 5.6. $\square$

Let be $w\in\mathcal{L}$ with $|w|\geq 2$. Define

$$\hat{l}(w) := -\log \sum_{w'\in\partial C(w)} L(o,w'|1).$$

We obtain the following law of large numbers:

Proposition 5.8.

$$\lim_{k\to\infty}\frac{\hat{l}(X_{\mathbf{e}_k})}{k} = H(\mathbf{Y}) \quad \text{almost surely.}$$

Proof. Let be $k\in\mathbb{N}$ and assume for the moment that $\mathbf{W}_l=y_la_lb_l$, where $y_l\in\mathcal{A}^*\setminus\{o\}$ and $a_lb_l\in\mathcal{A}^2$ for $0\leq l\leq k$. That is, $X_{\mathbf{e}_l}=y_0y_1\dots y_la_lb_l$. Furthermore, assume that $\mathbf{Y}_1=(j,t^{(1)})$, where $j=\tau(C(a_1b_1))$, and $\mathbf{Y}_l=(s^{(l)},t^{(l)})$ for $2\leq l\leq k$, where the values of $s^{(2)},\dots,s^{(k-1)}$ and $t^{(1)},\dots,t^{(k-1)}$ are determined by the values of $\mathbf{W}_l=y_la_lb_l$.

One can show that, for almost every realisation $(x_1, \underline{y}_1, \underline{y}_2, \dots)$ of $(X_{\mathbf{e}_1}, \mathbf{Y}_1, \mathbf{Y}_2, \dots)$,

$$H(\mathbf{Y}) = \lim_{k\to\infty} -\frac{1}{k} \log \mathbb{P}\bigl[C(X_{\mathbf{e}_1}) = C(x_1), \mathbf{Y}_1 = \underline{y}_1, \dots, \mathbf{Y}_k = \underline{y}_k\bigr]. \tag{5.7}$$

This follows from the fact that there are only finitely many possibilities for $C(X_{\mathbf{e}_1})$ which do not affect the resulting limit. Since the proof of this equation consists of technical reformulations of the involved probabilities we omit it at this place and give it in Lemma C.1 in Appendix C.

Recall from Equation (3.1) that $G(o,w|1) = G(o,o|1)L(o,w|1)$ for all $w \in \mathcal{L}$ and that $\xi(\cdot)$ can only take finitely many (non-zero) values. We now can conclude as follows:

$$\begin{aligned}
&\lim_{k\to\infty} \frac{\hat{l}(X_{\mathbf{e}_k})}{k} = \lim_{k\to\infty} -\frac{1}{k} \log \sum_{w' \in \partial C(y_0 y_1 \dots y_k a_k b_k)} L(o,w'|1) \\
= \;& \lim_{k\to\infty} -\frac{1}{k} \log \sum_{bc \in \mathcal{A}^2 : bc \in \partial C(a_k b_k)} L(o, y_0 y_1 \dots y_k bc|1) \\
= \;& \lim_{k\to\infty} -\frac{1}{k} \log \Biggl[\sum_{w_1 \in \partial C(y_0 y_1 a_1 b_1)} \sum_{\substack{w_2,\dots,w_k \in \mathcal{W}_0: \\ w_i \in \partial C(y_i a_i b_i) \\ \text{for all } 2 \le i \le k}} \sum_{\substack{w' \in \mathcal{L}: \\ w' \notin C(w_1)}} L(o,w'|1) p(w',w_1) \prod_{i=2}^{k} \mathbb{L}([w_{i-1}], w_i) \Biggr] \\
= \;& \lim_{k\to\infty} -\frac{1}{k} \log \Biggl[\sum_{\substack{w_1 \in \partial C(y_0 y_1 a_1 b_1); \\ w_2,\dots,w_k \in \mathcal{W}_0: \\ w_i \in \partial C(y_i a_i b_i) \\ \text{for all } 2 \le i \le k; \\ w' \in \mathcal{L} \setminus C(w_1)}} G(o,w'|1) p(w',w_1) \xi([w_1]) \cdot \prod_{i=2}^{k} \frac{\xi([w_i])}{\xi([w_{i-1}])} \mathbb{L}([w_{i-1}], w_i) \Biggr] \\
= \;& \lim_{k\to\infty} -\frac{1}{k} \log \Biggl[\sum_{w_1 \in \partial C(y_0 y_1 a_1 b_1)} \sum_{\substack{w_2,\dots,w_k \in \mathcal{W}_0: \\ w_i \in \partial C(y_i a_i b_i) \\ \text{for all } 2 \le i \le k}} \mathbb{P}[X_{\mathbf{e}_1} = w_1] q(y_1[w_1], w_2) \prod_{i=3}^{k} q(w_{i-1}, w_i) \Biggr] \\
= \;& \lim_{k\to\infty} -\frac{1}{k} \log \mathbb{P}\left[\begin{array}{c} X_{\mathbf{e}_1} \in C(y_0 y_1 a_1 b_1), \mathbf{Y}_1 = (j, t^{(1)}), \\ \mathbf{Y}_2 = (s^{(2)}, t^{(2)}), \dots, \mathbf{Y}_{k-1} = (s^{(k-1)}, t^{(k-1)}) \end{array} \right] = H(\mathbf{Y}).
\end{aligned}$$

The last equation follows from (5.7). We remark that the first coordinate of $\mathbf{Y}_1$ describes only the cone type of $X_{\mathbf{e}_1}$ but there may be several distinct cones of the same type $j \in \mathcal{I}$ with $j = \tau(C(X_{\mathbf{e}_1}))$. □

Recall the definition of $l(w) = -\log L(o,w|1)$ for $w \in \mathcal{L}$.

Corollary 5.9.

$$\lim_{k\to\infty} \frac{l(X_{\mathbf{e}_k})}{k} = H(\mathbf{Y}) \quad \text{almost surely.}$$

Proof. It suffices to compare $\hat{l}(X_{\mathbf{e}_k})$ with $l(X_{\mathbf{e}_k})$. Assume for a moment that $X_{\mathbf{e}_k} = w_k$ with $w_k \in \mathcal{L}$ and that $X_{\mathbf{e}_k}$ is on the boundary of the cone C_k. Then, the probability of walking *inside* C_k from any $w' \in \partial C_k$ to any $w - k \in \partial C_k$ (or vice versa) can be bounded from below by some constant ε_0, because the probabilities depend only on $[w_k], [w'] \in \mathcal{A}^2$: that is,

$$\mathbb{P}_{w'}[\exists n \in \mathbb{N} : X_n = w_k, \forall m \le n : X_n \in C(w')] \ge \varepsilon_0.$$

Therefore,

$$\begin{aligned} L(o, X_{\mathbf{e}_k}|1) &\leq \sum_{w' \in \partial C_k} L(o, w'|1) = \hat{l}(X_{\mathbf{e}_k}), \\ \hat{l}(X_{\mathbf{e}_k}) \cdot \varepsilon_0 &\leq \sum_{w' \in \partial C_k} L(o, w'|1) \cdot \mathbb{P}_{w'}[\exists n \in \mathbb{N} : X_n = w_k, \forall m \leq n : X_n \in C(w')] \\ &\leq |\mathcal{A}^2| \cdot L(o, X_{\mathbf{e}_k}|1). \end{aligned}$$

In the second inequality chain we extended paths from o to w' to paths from o to w_k via w' such that each such path is counted at most $|\mathcal{A}^2|$ times. Taking logarithms, dividing by k and letting k tend to infinity yields the claim. □

Now we come to an important law of large numbers. Denote by ν_0 the invariant probabilty measure of the positive recurrent Markov chain $(\mathbf{W}_k)_{k\in\mathbb{N}}$ and define

$$\lambda := \mathbb{E}[|\mathbf{W}_1^{(\nu)}|] - 2 = \sum_{w \in \mathcal{W}_0} \nu_0(w) \cdot \big(|w| - 2\big). \tag{5.8}$$

Then:

Proposition 5.10.

$$\lim_{k\to\infty} \frac{l(X_n)}{n} = \ell \cdot \lambda^{-1} \cdot H(\mathbf{Y}) \quad \text{almost surely.}$$

Proof. Define

$$\hat{\mathbf{e}}_k := \inf\big\{m \in \mathbb{N} \big| \forall n \geq m : |X_n| = k\big\}.$$

Observe that $\hat{\mathbf{e}}_k - 1 = \sup\big\{m \in \mathbb{N} \big| |X_m| = k-1\big\}$. Transience yields $\hat{\mathbf{e}}_k < \infty$ almost surely for all $k \in \mathbb{N}$. By [7, Proposition 2.3], $k/(\hat{\mathbf{e}}_k - 1)$ tends to the rate of escape ℓ as $k \to \infty$; hence, $k/\hat{\mathbf{e}}_k \to \ell$ as $k \to \infty$. Define the *maximal last entry times* at time $n \in \mathbb{N}$ as

$$\begin{aligned} \mathbf{k}(n) &:= \max\{k \in \mathbb{N} \mid \hat{\mathbf{e}}_k \leq n\}, \\ \mathbf{t}(n) &:= \max\{k \in \mathbb{N} \mid \mathbf{e}_k \leq n\}. \end{aligned}$$

Obviously, $\mathbf{k}(n) \geq \mathbf{t}(n)$ and each last entry time $\mathbf{e}_k$ corresponds (depending on the concrete realization) to exactly one $\hat{\mathbf{e}}_l$ with $l \geq k$. First, we rewrite

$$\frac{l(X_n)}{n} = \frac{l(X_n) - l(X_{\mathbf{e}_{\mathbf{t}(n)}})}{n} + \frac{l(X_{\mathbf{e}_{\mathbf{t}(n)}})}{\mathbf{t}(n)} \cdot \frac{\mathbf{t}(n)}{\mathbf{k}(n)} \cdot \frac{\mathbf{k}(n)}{\hat{\mathbf{e}}_{\mathbf{k}(n)}} \cdot \frac{\hat{\mathbf{e}}_{\mathbf{k}(n)}}{n}. \tag{5.9}$$

Let ε_1 be the minimal occuring positive single-step transition probability. Define

$$D := \max\left\{ |w_2| - |w_1| \,\middle|\, \begin{array}{c} \exists ab \in \mathcal{A}^2 : C(ab) \text{ has covering } C_1, \ldots, C_{n(ab)}, \\ w_1 \in \partial C(ab), w_2 \in \bigcup_{i=1}^{n(ab)} \partial C_i \end{array} \right\} < \infty.$$

Then we have $\hat{\mathbf{e}}_{\mathbf{k}(n)} \geq \mathbf{e}_{\mathbf{t}(n)} \geq \hat{\mathbf{e}}_{\mathbf{k}(n)-D}$ and $n/\mathbf{e}_{\mathbf{t}(n)} \geq 1$. This implies

$$1 \leq \frac{n}{\mathbf{e}_{\mathbf{t}(n)}} \leq \frac{\hat{\mathbf{e}}_{\mathbf{k}(n)+1}}{\hat{\mathbf{e}}_{\mathbf{k}(n)-D}} = \frac{\hat{\mathbf{e}}_{\mathbf{k}(n)+1}}{\mathbf{k}(n)} \frac{\mathbf{k}(n) - D}{\hat{\mathbf{e}}_{\mathbf{k}(n)-D}} \xrightarrow{n\to\infty} \frac{1}{\ell} \cdot \ell = 1 \quad \text{a.s.,} \tag{5.10}$$

which in turn yields $(n - \mathbf{e}_{\mathbf{t}(n)})/n \to 0$ as $n \to \infty$. Thus, the first quotient on the right hand side of (5.9) tends to zero since

$$\begin{aligned} L(o, X_n|1) \cdot \varepsilon_1^{n-\mathbf{e}_{\mathbf{t}(n)}} &\leq L(o, X_{\mathbf{e}_{\mathbf{t}(n)}}|1) \quad \text{(due to weak symmetry)}, \\ L(o, X_{\mathbf{e}_{\mathbf{t}(n)}}|1) \cdot \varepsilon_1^{n-\mathbf{e}_{\mathbf{t}(n)}} &\leq L(o, X_n|1). \end{aligned}$$

Here we used the fact that one can walk from $X_{\mathbf{e}_{\mathbf{t}(n)}}$ to X_n (or vice versa) in $n-\mathbf{e}_{\mathbf{t}(n)}$ steps. By Corollary 5.9, $l(X_{\mathbf{e}_{\mathbf{t}(n)}})/\mathbf{t}(n)$ tends to $H(\mathbf{Y})$. On the other hand side, $\hat{\mathbf{e}}_k/k$ tends almost surely to $1/\ell$ and $\hat{\mathbf{e}}_{\mathbf{k}(n)}/n$ tends to 1 almost surely since $1\leq n/\hat{\mathbf{e}}_{\mathbf{k}(n)}\leq n/\mathbf{e}_{\mathbf{t}(n)}\to 1$ by (5.10). It remains to investigate the limit $\lim_{k\to\infty}\mathbf{k}(n)/\mathbf{t}(n)$. Clearly,

$$\frac{\mathbf{k}(n)}{\mathbf{t}(n)}=\frac{|X_{\hat{\mathbf{e}}_{\mathbf{k}(n)}}|}{\mathbf{t}(n)}=\frac{1}{\mathbf{t}(n)}\Big(|X_{\mathbf{e}_1}|+\sum_{i=1}^{\mathbf{t}(n)-1}\big(|X_{\mathbf{e}_{i+1}}|-|X_{\mathbf{e}_i}|\big)+\big(|X_{\hat{\mathbf{e}}_{\mathbf{k}(n)}}|-|X_{\mathbf{e}_{\mathbf{t}(n)}}|\big)\Big).$$

Note that $0\leq |X_{\hat{\mathbf{e}}_{\mathbf{k}(n)}}|-|X_{\mathbf{e}_{\mathbf{t}(n)}}|\leq D$ and $0<|X_{\mathbf{e}_1}|\leq D_1$ almost surely for some suitable constant D_1. Thus, it is sufficient to consider

$$\frac{1}{k}\sum_{i=1}^{k}\big(|X_{\mathbf{e}_{i+1}}|-|X_{\mathbf{e}_i}|\big)=\frac{1}{k}\sum_{i=1}^{k}\big(|\mathbf{W}_i|-2\big).$$

Since $(\mathbf{W}_k)_{k\in\mathbb{N}}$ is positive recurrent, the ergodic theorem yields almost surely

$$\lim_{k\to\infty}\frac{1}{k}\sum_{i=1}^{k}\big(|\mathbf{W}_i|-2\big)=\sum_{w\in\mathcal{W}_0}\nu_0(w)\big(|w|-2\big)=\lambda.$$

This finishes the proof and gives the proposed formula. □

6 Existence of entropy

We now link Proposition 5.10 with the asymptotic entropy of the random walk $(X_n)_{n\in\mathbb{N}_0}$. For this purpose, we follow the reasoning of [8]. First, we need the following lemma:

Lemma 6.1. There is $R>1$ such that $G(w_1,w_2|R)<\infty$ for all $w_1,w_2\in\mathcal{L}$.

Proof. A simple adaption of the proof of [16, Proposition 8.2] shows that, for $w_1,w_2\in\mathcal{L}$, $G(w_1,w_2|z)$ has radius of convergence $R(w_1,w_2)>1$. At this point we also need the suffix-irreducibility Assumption 2.4; see Subsection A.1 for a comment on how to weaken this assumption. Since we assume the random walk $(X_n)_{n\in\mathbb{N}_0}$ to be irreducible, the radius of convergence is independent from w_1 and w_2; hence, $G(w_1,w_2|R)<\infty$ for all $w_1,w_2\in\mathcal{L}$ and $R=R(w_1,w_2)$. □

Let us remark that we have also $\bar{L}(ab,cde|R)<\infty$, $\overline{G}(ab,cd|R)<\infty$ and $L(o,a|R)<\infty$ for all $a,b,c,d,e\in\mathcal{A}$, since these generating functions are dominanted by Green functions. In the following let be $\varrho\in[1,R)$.

Lemma 6.2. There are constants D_1 and $D_2>0$ such that for all $m,n\in\mathbb{N}_0$

$$p^{(m)}(o,X_n)\leq D_1\cdot D_2^n\cdot\varrho^{-m}.$$

Proof. Denote by $\mathcal{C}_\varrho$ the circle with radius ϱ in the complex plane centered at 0. A straightforward computation together with Fubini's Theorem shows for $m\in\mathbb{N}_0$ and $w\in\mathcal{L}$:

$$\frac{1}{2\pi i}\oint_{\mathcal{C}_\varrho}G(o,w|z)\,z^{-m}\frac{dz}{z}=p^{(m)}(o,w);$$

compare with [8, Lemma 3.4]. Since $G(o,w|z)$ is analytic on $\mathcal{C}_\varrho$, we have $|G(o,w|z)|\leq G(o,w|\varrho)$ for all $|z|=\varrho$. Thus,

$$p^{(m)}(o,w)\leq\frac{1}{2\pi}\cdot\varrho^{-m-1}\cdot G(o,w|\varrho)\cdot 2\pi\varrho=G(o,w|\varrho)\cdot\varrho^{-m}.$$

Set $L := 1 \vee \max\{\bar{L}(ab, cde|\varrho) \mid a,b,c,d,e \in \mathcal{A}\}$, $C_0 := \varrho \cdot G(o,o|\varrho) \cdot \sum_{a\in\mathcal{A}} L(o,a|\varrho)$ and $C_1 = \max\{\overline{G}(ab,cd|\varrho) \mid ab, cd \in \mathcal{A}^2\}$. Equation (3.5) provides for all $w \in \mathcal{L}$ with $|w| \geq 2$

$$G(o,w|\varrho) = G(o,o|\varrho) \cdot L(o,w|\varrho) \leq C_0 \cdot |\mathcal{A}|^{2(|w|-2)} \cdot L^{|w|-2} \cdot C_1.$$

Set $C_2 := C_0 \vee \max\{G(o,w|\varrho) | w \in \mathcal{L}, |w| \leq 2\}$. Since $|X_n| \leq n$, we obtain the proposed inequality by setting $D_1 := C_1 + C_2$ and $D_2 := |\mathcal{A}|^2 \cdot L$:

$$p^{(m)}(o, X_n) \leq D_1 \cdot |\mathcal{A}|^{2|X_n|} \cdot L^{|X_n|} \cdot \varrho^{-m} \leq D_1 \cdot |\mathcal{A}|^{2n} \cdot L^n \cdot \varrho^{-m} = D_1 \cdot D_2^n \cdot \varrho^{-m}.$$

□

The following technical lemma will be used in the proof of the next theorem:

Lemma 6.3. Let $(A_n)_{n\in\mathbb{N}}$, $(a_n)_{n\in\mathbb{N}}$, $(b_n)_{n\in\mathbb{N}}$ be sequences of strictly positive numbers with $A_n = a_n + b_n$. Assume that $\lim_{n\to\infty} -\frac{1}{n}\log A_n = c \in [0,\infty)$ and that $\lim_{n\to\infty} b_n/q^n = 0$ for all $q \in (0,1)$. Then $\lim_{n\to\infty} -\frac{1}{n}\log a_n = c$.

Proof. A proof can be found in [8, Lemma 3.5]. □

Lemma 6.4. For $n \in \mathbb{N}$, consider the function $f_n : \mathcal{L} \to \mathbb{R}$ defined by

$$f_n(w) := \begin{cases} -\frac{1}{n}\log \sum_{m=0}^{n^2} p^{(m)}(o,w), & \text{if } p^{(n)}(o,w) > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Then there are constants d and D such that $d \leq f_n(w) \leq D$ for all $n \in \mathbb{N}$ and $w \in \mathcal{L}$.

Proof. Let be $w \in \mathcal{L}$ and $n \in \mathbb{N}$ with $p^{(n)}(o,w) > 0$. For $w_1 \in \mathcal{L}$ and $z > 0$, define the first return generating function as

$$U(w_1,w_1|z) := \sum_{n\geq 1} \mathbb{P}\big[X_n = w_1, \forall m \in \{1,\dots,n-1\} : X_m \neq w_1 \big| X_0 = w_1\big] \cdot z^n.$$

Recall the number $R > 1$ from Lemma 6.1. Then

$$G(w,w|1) \leq \frac{1}{1-\frac{1}{R}}; \tag{6.1}$$

indeed, since $G(w,w|z) = \big(1 - U(w,w|z)\big)^{-1}$ it must be that $U(w,w|z) < 1$ for all $w \in \mathcal{L}$ and all $z \in [0,R)$; moreover, $U(w,w|0) = 0$, $U(w,w|z)$ is continuous, strictly increasing and strictly convex for $z \in [0,R)$, so we must have $U(w,w|z) \leq 1/R$ for all $z \in [0,R)$, providing (6.1).

Define $F(o,w) := \sum_{n\geq 0} f^{(k)}(o,w)$, where $f^{(k)}(o,w)$ is the probability of starting at o and with the first visit to w at time k. By conditioning on the first visit to w we get $G(o,w|1) = F(o,w)G(w,w|1)$. Therefore,

$$\sum_{m=0}^{n^2} p^{(m)}(o,w) \leq G(o,w|1) = F(o,w) \cdot G(w,w|1) \leq \frac{1}{1-\frac{1}{R}},$$

that is,

$$f_n(w) \geq -\frac{1}{n}\log\frac{1}{1-\frac{1}{R}} \geq -\log\frac{1}{1-\frac{1}{R}} =: d.$$

For the upper bound, observe that $w \in \mathcal{L}$ with $p^{(n)}(o,w) > 0$ can be reached from o in n steps with a probability of at least ε_0^n, where

$$\varepsilon_0 := \min\{p(w_1,w_2) \mid w_1,w_2 \in \mathcal{A}^*, p(w_1,w_2) > 0\} > 0$$

is independent from w. Thus, the sum $\sum_{m=0}^{n^2} p^{(m)}(o,w)$ has a value greater or equal to ε_0^n. Hence, $f_n(x) \leq -\log\varepsilon_0 =: D$. □

Now we can finally prove:

Proof of Theorem 2.5. Recall Equation (3.1). We can rewrite $\ell \cdot \lambda^{-1} \cdot H(\mathbf{Y})$ as

$$\frac{\ell \cdot H(\mathbf{Y})}{\lambda} = \int \frac{\ell \cdot H(\mathbf{Y})}{\lambda}\, d\mathbb{P} = \int \lim_{n\to\infty} -\frac{1}{n} \log L\big(o, X_n(\omega)\big|1\big)\, d\mathbb{P}(\omega)$$
$$= \int \lim_{n\to\infty} -\frac{1}{n} \log \frac{G\big(o, X_n(\omega)\big|1\big)}{G(o,o|1)}\, d\mathbb{P}(\omega) = \int \lim_{n\to\infty} -\frac{1}{n} \log G\big(o, X_n(\omega)|1\big)\, d\mathbb{P}(\omega).$$

Recall that π_n denotes the distribution of X_n. Since

$$G(o, X_n|1) = \sum_{m\geq 0} p^{(m)}(o, X_n) \geq p^{(n)}(o, X_n) = \pi_n(X_n),$$

we have

$$\frac{\ell \cdot H(\mathbf{Y})}{\lambda} \leq \int \liminf_{n\to\infty} -\frac{1}{n} \log \pi_n\big(X_n(\omega)\big)\, d\mathbb{P}(\omega). \tag{6.2}$$

The next aim is to prove that $\limsup_{n\to\infty} -\frac{1}{n}\mathbb{E}\big[\log \pi_n(X_n)\big] \leq \ell \cdot H(\mathbf{Y})/\lambda$. We now apply Lemma 6.3 by setting

$$A_n := \sum_{m\geq 0} p^{(m)}(o, X_n),\ a_n := \sum_{m=0}^{n^2} p^{(m)}(o, X_n) \text{ and } b_n := \sum_{m\geq n^2+1} p^{(m)}(o, X_n).$$

By Lemma 6.2,

$$b_n \leq \sum_{m\geq n^2+1} D_1 \cdot D_2^n \cdot \varrho^{-m} = D_1 \cdot D_2^n \cdot \frac{\varrho^{-n^2-1}}{1-\varrho^{-1}}.$$

Therefore, b_n decays faster than any geometric sequence. Applying Lemma 6.3 together with (3.1) gives almost surely

$$\frac{\ell \cdot H(\mathbf{Y})}{\lambda} = \lim_{n\to\infty} -\frac{1}{n} \log L(o, X_n) = \lim_{n\to\infty} -\frac{1}{n} \log G(o, X_n) = \lim_{n\to\infty} -\frac{1}{n} \log \sum_{m=0}^{n^2} p^{(m)}\big(o, X_n\big).$$

Due to Lemma 6.4 we can apply the Dominated Convergence Theorem and get:

$$\frac{\ell \cdot H(\mathbf{Y})}{\lambda} = \int \frac{\ell \cdot H(\mathbf{Y})}{\lambda}\, d\mathbb{P} = \int \lim_{n\to\infty} -\frac{1}{n} \log \sum_{m=0}^{n^2} p^{(m)}(o, X_n)\, d\mathbb{P}$$
$$= \lim_{n\to\infty} \int -\frac{1}{n} \log \sum_{m=0}^{n^2} p^{(m)}(o, X_n)\, d\mathbb{P} = \lim_{n\to\infty} -\frac{1}{n} \sum_{w\in\mathcal{L}} p^{(n)}(o, w) \log \sum_{m=0}^{n^2} p^{(m)}(o, w).$$

For $w \in \mathcal{L}$, define the following distribution μ_0 on $\mathcal{L}$:

$$\mu_0(w) := \frac{1}{n^2+1} \sum_{m=0}^{n^2} p^{(m)}(o, w).$$

Recall that the non-negativity of the Kullback-Leibler divergence (in this context also called *Shannon's Inequality*) gives

$$-\sum_{w\in\mathcal{L}} p^{(n)}(o, w) \log \mu_0(w) \geq -\sum_{w\in\mathcal{L}} p^{(n)}(o, w) \log p^{(n)}(o, w).$$

Therefore,

$$\begin{aligned}\frac{\ell\cdot H(\mathbf{Y})}{\lambda} &\geq \limsup_{n\to\infty} -\frac{1}{n}\sum_{w\in\mathcal{L}} p^{(n)}(o,w)\log(n^2+1) - \frac{1}{n}\sum_{w\in\mathcal{L}} p^{(n)}(o,w)\log p^{(n)}(o,w)\\ &= \limsup_{n\to\infty} -\frac{1}{n}\int \log \pi_n(X_n)\,d\mathbb{P}.\end{aligned}$$

Now we can conclude with (6.2) and Fatou's Lemma:

$$\begin{aligned}\frac{\ell\cdot H(\mathbf{Y})}{\lambda} &\leq \int \liminf_{n\to\infty} -\frac{1}{n}\log\pi_n(X_n)d\mathbb{P} \leq \liminf_{n\to\infty}\int -\frac{1}{n}\log\pi_n(X_n)d\mathbb{P}\\ &\leq \limsup_{n\to\infty}\int -\frac{1}{n}\log\pi_n(X_n)d\mathbb{P} \leq \frac{\ell\cdot H(\mathbf{Y})}{\lambda}.\end{aligned}$$

Thus, the asymptotic entropy $h := \lim_{n\to\infty} -\frac{1}{n}\mathbb{E}\big[\log\pi_n(X_n)\big]$ exists and equals $\ell\cdot H(\mathbf{Y})/\lambda$. □

Finally, we can prove:

Proof of Corollary 2.7. The proofs of the statements in Corollary 2.7 are completely analogous to the proofs in [8, Corollary 3.9, Lemma 3.10], where [8, Lemma 3.10] holds also in the case $h=0$. □

Proof of Corollary 2.8. Recall the definition of $F(o,w)$ from the proof of Lemma 6.4 and the equation $G(o,w|1)=F(o,w)G(w,w|1)$. This yields together with (3.1):

$$\mathbb{P}[\exists n\in\mathbb{N}_0 : X_n = w] = F(o,w) = \frac{G(o,w|1)}{G(w,w|1)} = \frac{G(o,o|1)}{G(w,w|1)}L(o,w|1).$$

Since $1\leq G(X_n,X_n|1)\leq 1/(1-\frac{1}{R})$ with R from Lemma 6.1, we obtain the proposed result due to Proposition 5.10. □

7 Calculation of the entropy

In this section we collect several results about the asymptotic entropy. We show how the entropy can be calculated numerically or even exactly in some special cases, and we give some inequalities.

7.1 Numerical calculation and inequalities

In order to compute $h=\ell\cdot H(\mathbf{Y})/\lambda$ we have to calculate the three factors: while there are formulas for ℓ (see [7, Theorem 2.4]) and λ (given by (5.8)), it remains to explain how to calculate $H(\mathbf{Y})$. For this purpose, define for random variables $A_1,\dots,A_n$ on a finite state space $\mathcal{W}_A$ the *joint entropy* as

$$H(A_1,\dots,A_n) := -\sum_{a_1,\dots,a_n\in\mathcal{W}_A}\mathbb{P}\big[A_1=a_1,\dots,A_n=a_n\big]\log\mathbb{P}\big[A_1=a_1,\dots,A_n=a_n\big],$$

and let the *conditional entropy* $H(A_n|A_1,\dots,A_{n-1})$ be defined as

$$-\sum_{a_1,\dots,a_n\in\mathcal{W}_A}\mathbb{P}\big[A_1=a_1,\dots,A_n=a_n\big]\log\mathbb{P}\big[A_n=a_n\big|A_1=a_1,\dots,A_{n-1}=a_{n-1}\big].$$

Here, we set $0\cdot\log 0 := 0$, since $x\log x\to 0$ as $x\to 0+$. By Cover and Thomas [5, Theorem 4.2.1], we have $H(\mathbf{Y}) = \lim_{n\to\infty}\frac{1}{n}H(\mathbf{Y}_1^{(\nu)},\dots,\mathbf{Y}_n^{(\nu)})$. In general, the computation of $H(\mathbf{Y})$ is a hard task. But there is a simple way for a numerical calculation of $H(\mathbf{Y})$, which follows from the inequalities

$$H\big(\mathbf{Y}_n^{(\nu)}\big|\big((\mathbf{i}_1^{(\nu)},\mathbf{W}_1^{(\nu)}),(\mathbf{i}_2^{(\nu)},\mathbf{W}_2^{(\nu)})\big),\mathbf{Y}_1^{(\nu)},\dots,\mathbf{Y}_{n-1}^{(\nu)}\big) \leq H(\mathbf{Y}) \leq H(\mathbf{Y}_n^{(\nu)}\mid \mathbf{Y}_1^{(\nu)},\dots,\mathbf{Y}_{n-1}^{(\nu)}) \tag{7.1}$$

for all $n \in \mathbb{N}$; see [5, Theorem 4.5.1]. In particular, it is even shown that

$$H(\mathbf{Y}_n^{(\nu)} \mid \mathbf{Y}_1^{(\nu)}, \dots, \mathbf{Y}_{n-1}^{(\nu)}) - H\big(\mathbf{Y}_n^{(\nu)} \big| \big((\mathbf{i}_1^{(\nu)}, \mathbf{W}_1^{(\nu)}), (\mathbf{i}_2^{(\nu)}, \mathbf{W}_2^{(\nu)})\big), \mathbf{Y}_1^{(\nu)}, \dots, \mathbf{Y}_{n-1}^{(\nu)}\big) \xrightarrow{n\to\infty} 0.$$

Hence, one can calculate $H(\mathbf{Y})$ numerically up to an arbitrarily small error. Obviously, this numerical approach depends strongly on the ability to solve the system of equations given by (3.2).

We now investigate whether the entropy is non-zero or not.

Corollary 7.1. If the random walk is expanding, then $h > 0$. Otherwise, $h = 0$.

Proof. Take any $(i_{k,l}, w_1), (j_{p,q}, w_2) \in \mathcal{W}$ with

$$\mathbb{P}[(\mathbf{i}_1^{(\nu)}, \mathbf{W}_1^{(\nu)}) = (i_{k,l}, w_1), (\mathbf{i}_2^{(\nu)}, \mathbf{W}_2^{(\nu)}) = (j_{p,q}, w_2)] > 0.$$

The values $(i_{k,l}, w_1), (j_{p,q}, w_2)$ determine the value of $\mathbf{Y}_1^{(\nu)}$ uniquely. In the expanding case, there are at least two elements $(s_{j,m}, w'), (t_{j,n}, w'') \in \mathcal{W}$ such that $w', w'' \in C([w_2])$ with $C(w') \cap C(w'') = \emptyset$ and $q(w_2, w') > 0$ and $q(w_2, w'') > 0$, yielding $\pi\big((j_{p,q}, w_2), (s_{j,m}, w')\big) \neq \pi\big((j_{p,q}, w_2), (t_{j,n}, w'')\big)$. Let w' be in the m-th cone of type s in the covering of $C([w_2])$. Then set

$$\begin{aligned} & P\big((i_{k,l}, w_1), (j_{p,q}, w_2), (s_{j,m}, w')\big) \\ := \ & \mathbb{P}\big[\mathbf{Y}_2^{(\nu)} = \big(\tau(C(w_2)), s_m\big) \,\big|\, (\mathbf{i}_1^{(\nu)}, \mathbf{W}_1^{(\nu)}) = (i_{k,l}, w_1), (\mathbf{i}_2^{(\nu)}, \mathbf{W}_2^{(\nu)}) = (j_{p,q}, w_2)\big] \\ \geq \ & q(w_2, w') > 0. \end{aligned}$$

We also have $P\big((i_{k,l}, w_1), (j_{p,q}, w_2), (t_{j,n}, w'')\big) > 0$ since $q(w_2, w'') > 0$ and $C(w') \cap C(w'') = \emptyset$. This implies $P\big((i_{k,l}, w_1), (j_{p,q}, w_2), (s_{j,m}, w')\big) < 1$. From (7.1) follows then

$$\begin{aligned} H(\mathbf{Y}) \ &\geq \ H\big(\mathbf{Y}_2^{(\nu)} \big| \big((\mathbf{i}_1^{(\nu)}, \mathbf{W}_1^{(\nu)}), (\mathbf{i}_2^{(\nu)}, \mathbf{W}_2^{(\nu)})\big), \mathbf{Y}_1^{(\nu)}\big) \\ &\geq \ P\big((i_{k,l}, w_1), (j_{p,q}, w_2), (s_{j,m}, w')\big) \log P\big((i_{k,l}, w_1), (j_{p,q}, w_2), (s_{j,m}, w')\big) > 0. \end{aligned}$$

Thus, we have shown that $h > 0$ if $(X_n)_{n\in\mathbb{N}_0}$ is expanding.

Now consider the case when the random walk on $\mathcal{L}$ is *not* expanding. Then each cone has a covering consisting of only one single subcone. This implies that the value $\tau(C(\mathbf{W}_1^{(\nu)})) = \mathbf{i}_1^{(\nu)}$ determines uniquely the values $\tau(C(\mathbf{W}_k^{(\nu)}))$ for $k \geq 2$. Moreover, given the value of $\tau(C(\mathbf{W}_1^{(\nu)}))$ the values of $\mathbf{Y}_k^{(\nu)}$, $k \geq 1$, are deterministic. That is, $\mathbf{Y}_n^{(\nu)}$ is uniquely determined by $\mathbf{Y}_1^{(\nu)}$, hence $\mathbb{P}[\mathbf{Y}_n^{(\nu)} = \cdot \mid \mathbf{Y}_1^{(\nu)} = (s, t_n)] \in \{0, 1\}$. This implies

$$0 \leq H(\mathbf{Y}) \leq H(\mathbf{Y}_n^{(\nu)} \mid \mathbf{Y}_1^{(\nu)}, \dots, \mathbf{Y}_{n-1}^{(\nu)}) \leq H(\mathbf{Y}_n^{(\nu)} \mid \mathbf{Y}_1^{(\nu)}) = 0,$$

where the last inequality follows from [5, Theorem 2.6.5]. Thus, $h = 0$. □

In order to get a complete picture, we show that the entropy is zero for recurrent random walks:

Corollary 7.2. If $(X_n)_{n\in\mathbb{N}_0}$ is recurrent then $h = 0$.

Proof. Clearly, $-\frac{1}{n}\mathbb{E}\big[\log \pi_n(X_n)\big] \geq 0$. Assume now that $\limsup_{n\to\infty} -\frac{1}{n}\mathbb{E}\big[\log \pi_n(X_n)\big] = c > 0$. Then there is a (deterministic) sequence $(n_k)_{k\in\mathbb{N}}$ such that, for any $\varepsilon_1 \in (0, c)$,

$$-\frac{1}{n_k}\mathbb{E}\big[\log \pi_{n_k}(X_{n_k})\big] \geq c - \varepsilon_1 > 0 \tag{7.2}$$

for all $k \in \mathbb{N}$. Denote by ε_0 the minimal occuring positive single-step transition probability of $(X_n)_{n\in\mathbb{N}_0}$. Then $-\frac{1}{n_k}\log \pi_{n_k}(X_{n_k}) \leq -\log \varepsilon_0$. Choose $N \in \mathbb{N}$ with $1/N < c - \varepsilon_1$. Then there is some $\delta > 0$ with

$$\mathbb{P}\Big[-\frac{1}{n_k}\log \pi_{n_k}(X_{n_k}) \geq \frac{1}{N}\Big] \geq \delta \quad \forall k \in \mathbb{N}.$$

To see this, assume that $\delta = \delta_k$ depends on k with $\liminf_{k\to\infty} \delta_k = 0$: then we get with (7.2)

$$(-\log \varepsilon_0)\cdot \delta_k + (1-\delta_k)\frac{1}{N} \geq -\frac{1}{n_k}\mathbb{E}\big[\log \pi_{n_k}(X_{n_k})\big] \geq c - \varepsilon_1;$$

If δ_k tends to zero then we get a contradiction to the choice of N.

Choose now $\varepsilon > 0$ arbitrarily small with $\varepsilon < \delta$. Since $\ell = 0$ in the recurrent case, there is some index $K \in \mathbb{N}$ such that for all $k \geq K$:

$$\delta - \varepsilon \leq \mathbb{P}\big[-\log \pi_{n_k}(X_{n_k}) \geq n_k/N, |X_{n_k}| \leq \varepsilon n_k\big] \leq e^{-n_k/N}\cdot |\mathcal{A}|^{\varepsilon n_k}$$

which yields the inequality

$$\frac{1}{N} + \frac{1}{n_k}\log(\delta-\varepsilon) \leq \varepsilon \log|\mathcal{A}|.$$

But this gives a contradiction if we make ε sufficiently small since the right hand side tends to zero, but the left hand side to $\frac{1}{N}$ as $k\to\infty$. Thus, $\limsup_{n\to\infty} -\frac{1}{n}\mathbb{E}\big[\log \pi_n(X_n)\big] = 0$, yielding $h = 0$. □

Finally, we state an inequality which connects entropy, drift and growth. For this purpose, define $\mathcal{A}^*_{\leq n} = \{w \in \mathcal{A}^* \mid |w| \leq n\}$ for $n > 0$. The *growth* of $\mathcal{A}^*$ is then given by $g := \lim_{n\to\infty} \frac{1}{n}\log|\mathcal{A}^*_{\leq n}|$. Since $|\mathcal{A}^n| \leq |\mathcal{A}^*_{\leq n}| \leq n|\mathcal{A}^n|$, we have $g = \log|\mathcal{A}|$. We get the following connection between entropy, drift and growth:

Theorem 7.3. $h \leq \ell \cdot \log|\mathcal{A}|$.

Proof. Let be $\varepsilon > 0$. By Corollary 2.7 (1), there is some $N_\varepsilon \in \mathbb{N}$ such that for all $n \geq N_\varepsilon$:

$$1-\varepsilon \leq \mathbb{P}\big[-\log \pi_n(X_n) \geq (h-\varepsilon)n, |X_n| \leq (\ell+\varepsilon)n\big] \leq e^{-(h-\varepsilon)n}\cdot |\mathcal{A}^*_{\leq(\ell+\varepsilon)n}|.$$

Taking logarithms and dividing by n gives

$$(h-\varepsilon) + \frac{1}{n}\log(1-\varepsilon) \leq (\ell+\varepsilon)\cdot \frac{1}{(\ell+\varepsilon)n}\log|\mathcal{A}^*_{\leq(\ell+\varepsilon)n}|.$$

Making ε arbirtraily small and sending $n\to\infty$ yields the proposed claim. □

Let us remark that similar inequalities have been proved by Kaimanovich and Woess [14] for time and space homogeneous random walks and in [8] for random walks on free products.

7.2 Exact formula for unambiguous cone boundaries

In this subsection we give an exact formula for the asymptotic entropy in some special case. We call $ab \in \mathcal{A}^2$ *unambiguous* if $\partial C(ab) = \{ab\}$. In other words, whenever the random walk enters a subcone of type $C(wab)$, $w \in \mathcal{A}^*$, it must enter it through its single boundary point wab. We call the cone type $\tau(C(ab))$ also unambiguous. Existence of an unambiguous cone allows us to "cut" the random walk into i.i.d. pieces and to obtain a formula for the entropy $H(\mathbf{Y})$. For $n \in \mathbb{N}$, $x_2,\dots,x_n \in \mathcal{W}_0$ and unambiguous $ab \in \mathcal{A}^2$ define

$$\begin{aligned}
w(ab,x_2,\dots,x_n) &:= \mathbb{P}\big[\mathbf{W}_2 = x_2,\dots,\mathbf{W}_n = x_n, [\mathbf{W}_n] = ab \big| [\mathbf{W}_1] = ab\big],\\
\tilde{w}(ab,x_2,\dots,x_n) &:= \sum_{\substack{y_2,\dots,y_n\in\mathcal{W}_0:\\ y_i\in\partial C(x_i)\\ \text{for } 2\leq i\leq n}} \mathbb{P}\big[\mathbf{W}_2 = y_2,\dots,\mathbf{W}_n = y_n, [\mathbf{W}_n] = ab \big| [\mathbf{W}_1] = ab\big],
\end{aligned}$$

In particular, $\tilde{w}(ab, x_2) = \mathbb{P}\big[\mathbf{W}_2 = x_2, [\mathbf{W}_2] = ab \big| [\mathbf{W}_1] = ab\big]$. Recall that ν denotes the invariant probability measure of the process $(\mathbf{i}_k, \mathbf{W}_k)_{k\in\mathbb{N}}$. For unambiguous $ab \in \mathcal{A}^2$, set

$$\nu_{ab} := \sum_{(i_{m,n},x)\in\mathcal{W}:[x]=ab} \nu(i_{m,n}, x).$$

Then:

Theorem 7.4. If $ab \in \mathcal{A}^2$ is unambiguous, then

$$H(\mathbf{Y}) = -\nu_{ab} \sum_{n\geq 1} \sum_{\substack{x_2,\dots x_{n-1}\in\mathcal{W}_0:\\ [x_i]\neq ab \text{ for } 2\leq i\leq n-1}} \sum_{\substack{x_n\in\mathcal{W}_0:\\ [x_n]=ab}} w(ab, x_2, \dots, x_n) \log \tilde{w}(ab, x_2, \dots, x_n).$$

Proof. Write $\alpha := \tau(C(ab))$. By Proposition 5.6, we have

$$-\frac{1}{n} \log \mathbb{P}[\mathbf{Y}_1 = \underline{y}_1, \dots, \mathbf{Y}_n = \underline{y}_n] \xrightarrow{n\to\infty} H(\mathbf{Y})$$

for almost every trajectory $(\underline{y}_1, \underline{y}_2, \dots) \in \mathcal{W}_\pi^{\mathbb{N}}$. For any such trajectory, we define

$$N_0 := \min\big\{m \in \mathbb{N} \big| \tau(\mathbf{W}_{m+1}) = \alpha\big\} \text{ and } N_k := \min\big\{m \in \mathbb{N} \big| m > N_{k-1}, \tau(\mathbf{W}_{m+1}) = \alpha\big\}.$$

Define $d(n) := \max\{k \in \mathbb{N}_0 \mid N_k \leq n\}$. Since $\mathbf{Y}_{N_j}$ has the form $(t, \alpha_{t^{(n)},m})$ for some cone type $t \in \mathcal{I}$, $1 \leq m \leq n(t, \alpha)$, and $[\mathbf{W}_{N_k+1}] = ab$ for all $k \in \mathbb{N}$ we can use the strong Markov property as follows for all $n \geq 1$ and almost every trajectory $(\underline{y}_1, \underline{y}_2, \dots) \in \mathcal{W}_\pi^{\mathbb{N}}$:

$$\begin{aligned} & \mathbb{P}\big[\mathbf{Y}_{N_j+1} = \underline{y}_{N_j+1}, \dots, \mathbf{Y}_{N_j+n} = \underline{y}_n \mid \mathbf{Y}_1 = \underline{y}_1, \dots, \mathbf{Y}_{N_j} = \underline{y}_{N_j}\big] \\ = \;& \mathbb{P}\big[\mathbf{Y}_{N_j+1} = \underline{y}_{N_j+1}, \dots, \mathbf{Y}_{N_j+n} = \underline{y}_n \mid [\mathbf{W}_{N_j+1}] = ab\big]. \end{aligned}$$

In other words, the $\mathbf{Y}_k$'s collect only the information which cones are entered successively, but we know that the $(N_j + 1)$-th cone is entered through a boundary point with last two letters ab; hence, one can restart the process at some word ending with ab in the above equation without changing probabilities. Therefore, we can rewrite the following probability $\mathbb{P}\big[\mathbf{Y}_1 = \underline{y}_1, \dots, \mathbf{Y}_{d(n)} = \underline{y}_{d(n)}\big]$ as

$$\mathbb{P}\big[\mathbf{Y}_1 = \underline{y}_1, \dots, \mathbf{Y}_{N_0} = \underline{y}_{N_0}\big] \prod_{i=0}^{d(n)-1} \mathbb{P}\big[\mathbf{Y}_{N_i+1} = \underline{y}_{N_i+1}, \dots, \mathbf{Y}_{N_{i+1}} = \underline{y}_{N_{i+1}} \big| [\mathbf{W}_{N_i+1}] = ab\big].$$

Observe that the terms $\log \mathbb{P}\big[\mathbf{Y}_{N_i+1} = \cdot, \dots, \mathbf{Y}_{N_{i+1}} = \cdot \big| [\mathbf{W}_{N_i+1}] = ab\big]$, $i \in \mathbb{N}$, are i.i.d., since one can think of starting at some $\mathbf{W}_k$ with $[\mathbf{W}_k] = ab$ and stopping at the first time $l > k$ with $[\mathbf{W}_l] = ab$. By the ergodic theorem for positive recurrent Markov chains, $d(n)/n$ tends almost surely to ν_{ab}. Hence, if we consider only the subsequence where n equals one of the N_k's we obtain the following convergence for almost every trajectory $(\underline{y}_1, \underline{y}_2, \dots) \in \mathcal{W}_\pi^{\mathbb{N}}$ by classical ergodic theory:

$$\begin{aligned} & -\frac{1}{n} \log \mathbb{P}\big[\mathbf{Y}_1 = \underline{y}_1, \dots, \mathbf{Y}_{d(n)} = \underline{y}_{d(n)}\big] \\ = \quad & -\frac{d(n)}{n} \frac{1}{d(n)} \Bigg[\log \mathbb{P}\big[\mathbf{Y}_1 = \underline{y}_1, \dots, \mathbf{Y}_{N_0} = \underline{y}_{N_0}\big] \\ & \quad + \sum_{i=0}^{d(n)-1} \log \mathbb{P}\big[\mathbf{Y}_{N_i+1} = \underline{y}_{N_i+1}, \dots, \mathbf{Y}_{N_{i+1}} = \underline{y}_{N_{i+1}} \big| [\mathbf{W}_{N_i+1}] = ab\big]\Bigg] \\ \xrightarrow{n\to\infty} \quad & -\nu_{ab} \sum_{k\geq 1} \sum_{\substack{x_2,\dots,x_{k-1}\in\mathcal{W}_0:\\ [x_i]\neq ab\\ \text{for } 2\leq i\leq k-1}} \sum_{\substack{x\in\mathcal{W}_0:\\ [x]=ab}} w(ab, x_2, \dots, x_{k-1}, x) \log \tilde{w}(ab, x_2, \dots, x_{k-1}, x). \end{aligned}$$

This proves the claim. □

8 Analyticity of entropy

The random walk on $\mathcal{A}^*$ depends on *finitely* many parameters which are described by the transition probabilities $p(w_1,w_2)$, $w_1,w_2\in\mathcal{A}^*$ with $|w_1|\leq 2$ and $|w_2|\leq 3$; see (2.1). That is, each random walk on $\mathcal{A}^*$ can be defined via a vector $\underline{p}\in\mathbb{R}_+^{|\mathcal{B}|}$, where

$$\mathcal{B}:=\Bigl\{(w_1,w_2)\,\Big|\,w_1\in\mathcal{A}\cup\mathcal{A}^2\cup\{o\}, w_2\in\bigcup_{n=1}^{3}\mathcal{A}^n\cup\{o\}, \bigl||w_1|-|w_2|\bigr|\leq 1\Bigr\}.$$

In other words, the entry of $\underline{p}$ associated with the index $(w_1,w_2)\in\mathcal{B}$ describes the value of $p(w_1,w_2)$. The support $\mathrm{supp}(\underline{p})$ of $\underline{p}$ is the set of indices in $\mathcal{B}$ corresponding to non-zero entries of $\underline{p}$. Fix now any $\underline{p}_0\in\mathbb{R}_+^{|\mathcal{B}|}$ such that $\underline{p}_0$ describes a well-defined, transient random walk on $\mathcal{A}^*$, and let $\mathcal{P}(\underline{p}_0)$ be the set of vectors $\underline{p}\in\mathbb{R}^{|\mathcal{B}|}$ with support $\mathrm{supp}(\underline{p}_0)$ which allow well-defined, transient random walks on $\mathcal{A}^*$. The set $\mathcal{P}(\underline{p}_0)$ can be described by an open polygonal bounded convex set in $\mathbb{R}^d$ with some suitable $d\leq|\mathcal{B}|-1$ which depends on $\mathrm{supp}(\underline{p}_0)$; recall that $\ell>0$ if and only if $(X_n)_{n\in\mathbb{N}_0}$ is transient, and from the formula of ℓ in [7, Theorem 2.4] follows that ℓ varies continuously in $\underline{p}$, yielding that there is some open neighbourhood of $\underline{p}_0$ in $\mathbb{R}^d$ where $(X_n)_{n\in\mathbb{N}_0}$ remains still transient. We now ask whether the entropy mapping $\underline{p}\mapsto h=h_{\underline{p}}$ varies real-analytically on $\mathcal{P}(\underline{p}_0)$.

In the next subsection we will introduce a new Markov chain which is related to the last entry time process and leads under the projection $\pi(\cdot,\cdot)$ to a hidden Markov chain with same distribution as $(\mathbf{Y}_k)_{k\in\mathbb{N}}$. Afterwards we will be able to prove Theorem 2.6 in Subsection 8.2.

8.1 Modified last entry time process

The aim of this subsection is the construction of a Markov chain related to the last entry time process $(\mathbf{i}_k,\mathbf{W}_k)_{k\in\mathbb{N}}$ such that the transition matrix has strictly positive entries and the modified process leads under $\pi(\cdot,\cdot)$ (see (5.5)) to a hidden Markov chain with same asymptotic entropy.

Let be $ab,a_1b_1,a_2b_2\in\mathcal{A}^2$, and let $C_{j_{i,1}}$ be the first cone of type j in the covering of $C(a_1b_1)$ with $\tau(C(a_1b_1))=i$ and let $C_{j_{k,l}}$ be the l-th subcone of type j in the covering of $C(a_2b_2)$ with $\tau(C(a_2b_2))=k$. Assume that $y_0\in\partial C_{j_{k,l}}$ with $[y_0]=ab$. Since $C_{j_{i,1}}$ and $C_{j_{k,l}}$ are isomorphic, there is some unique $\bar{y}_0^{[i,j,ab]}\in\mathcal{A}^*$ such that $\bar{y}_0^{[i,j,ab]}ab\in\partial C_{j_{i,1}}$; see Section 4.1. In the following we will sometimes omit the superindex $[i,j,ab]$ and use the notation $\bar{y}_0=\bar{y}_0^{[i,j,ab]}$ for describing this replacement.

For $i,j\in\mathcal{I}$ and $ab\in\mathcal{A}^2$ with $\tau(C(ab))=j$, we write

$$\#\{j_{s,t}\mid s\neq i,ab\}:=\Bigl|\bigl\{(j_{s,t},w)\in\mathcal{W}\bigm|[w]=ab, s\in\mathcal{I}\setminus\{i\}, 1\leq t\leq n(s,j)\bigr\}\Bigr|.$$

It is not hard to see that $\#\{j_{s,t}\mid s\neq i,a_1b_1\}=\#\{j_{s,t}\mid s\neq i,a_2b_2\}$ if $\tau(C(a_1b_1))=\tau(C(a_2b_2))$ but this will not be relevant for our proofs, so we omit further explanations. Let be $(i_{k,l},x),(j_{m,n},y)\in\mathcal{W}$ with $[y]=ab\in\mathcal{A}^2$. This implies $\tau(C(x))=i$ and $y^{[i,j,ab]}\in\partial C_{j_{i,1}}$, where $C_{j_{i,1}}$ is the first cone of type j in the covering of $C([x])$. Define the following transition probabilities on $\mathcal{W}$:

$$\hat{q}\bigl((i_{k,l},x),(j_{m,n},y)\bigr):=\begin{cases}\frac{1}{\#\{j_{s,t}\mid s\neq i,ab\}+1}\frac{\xi([y])}{\xi([x])}\mathbb{L}(x,y), & \text{if } m=i\wedge n=1,\\ \frac{\xi([y])}{\xi([x])}\mathbb{L}(x,y), & \text{if } m=i\wedge n\geq 2,\\ \frac{1}{\#\{j_{s,t}\mid s\neq i,ab\}+1}\frac{\xi([y])}{\xi([x])}\mathbb{L}(x,\bar{y}^{[i,j,ab]}ab), & \text{if } m\neq i.\end{cases}$$

It is easy to see that these transition probabilities define a Markov chain (inherited from the Markov chain $(\mathbf{i}_k,\mathbf{W}_k)_{k\in\mathbb{N}}$): in the case $m=i\wedge n\geq 2$ we just have

$$\hat{q}\bigl((i_{k,l},x),(j_{m,n},y)\bigr)=\mathbb{P}\bigl[(\mathbf{i}_2,\mathbf{W}_2)=(j_{m,n},y)\mid(\mathbf{i}_1,\mathbf{W}_1)=(i_{k,l},x)\bigr];$$

otherwise we have, for $(j_{i,1},y)\in\mathcal{W}$,

$$\begin{aligned}&\hat{q}\big((i_{k,l},x),(j_{i,1},y)\big)+\sum_{\substack{(j_{s,t},w)\in\mathcal{W}:\\ s\neq i,[w]=ab,\\ 1\leq t\leq n(s,j)}}\hat{q}\big((i_{k,l},x),(j_{s,t},w)\big)\\ =\;&\mathbb{P}\big[(\mathbf{i}_2,\mathbf{W}_2)=(j_{i,1},y)\mid(\mathbf{i}_2,\mathbf{W}_2)=(i_{k,l},x)\big]\end{aligned}$$

since $y=\bar{y}^{[i,j,ab]}ab$ by definition. In other words, each step from $(i_{k,l},x)$ to $(j_{m,n},y)$ either behaves according to (5.4) (case $m=i$ and $n\geq 2$) or the step from $(i_{k,l},x)$ to $(j_{i,1},y)$ (when seen as a step of the process $(\mathbf{i}_k,\mathbf{W}_k)_{k\in\mathbb{N}}$)) is split up into different equally likely steps $(i_{k,l},x)$ to $(j_{m,n},\bar{y}ab)$ with $m\neq i$ or $m=i\wedge n=1$. Observe that the transitions depend only on $[x]$ in the first argument of $\hat{q}(\cdot,\cdot)$. By Proposition 5.4, the transition matrix $\widehat{Q}=\big(\hat{q}((i_{k,l},x),(j_{m,n},y))\big)$ is stochastic and governs a positive recurrent, aperiodic Markov chain $(\mathbf{t}_k,\mathbf{x}_k)_{k\in\mathbb{N}}$. In particular, $\widehat{Q}$ has strictly positive entries. The initial distribution $\hat{\mu}_1$ of $(\mathbf{t}_1,\mathbf{x}_1)$ is defined as

$$\hat{\mu}_1(i_{m,n},x):=\mathbb{P}[(\mathbf{i}_1,\mathbf{W}_1)=(i_{m,n},x)]>0$$

for $(i_{m,n},x)\in\mathcal{W}$.

The process $\big((\mathbf{t}_k,\mathbf{x}_k),(\mathbf{t}_{k+1},\mathbf{x}_{k+1})\big)_{k\in\mathbb{N}}$ is again a positive recurrent, aperiodic Markov chain whose transition matrix is denoted by $\widehat{Q}_2$ (arising from $\widehat{Q}$). We now define a new hidden Markov chain $(\mathbf{Z}_k)_{k\in\mathbb{N}}$ by

$$\mathbf{Z}_k:=\pi\big((\mathbf{t}_k,\mathbf{x}_k),(\mathbf{t}_{k+1},\mathbf{x}_{k+1})\big).$$

Observe that at this point the second branch in the definition of π in (5.5) comes into play for the definition of $\mathbf{Z}_k$. The crucial point is the following proposition:

Proposition 8.1. For all $(s^{(1)},t^{(1)}),\dots,(s^{(n)},t^{(n)})\in\mathcal{W}_\pi$,

$$\mathbb{P}\big[\mathbf{Y}_1=(s^{(1)},t^{(1)}),\dots,\mathbf{Y}_n=(s^{(n)},t^{(n)})\big]=\mathbb{P}\big[\mathbf{Z}_1=(s^{(1)},t^{(1)}),\dots,\mathbf{Z}_n=(s^{(n)},t^{(n)})\big].$$

Since the proof of this proposition consists of an long induction with tedious calculations we omit it at this place and give it in Appendix C.

The statement of the last proposition can be formulated in other words: the process governed by $\widehat{Q}$ can be seen as a last entry time process, where one has more subcones to enter (namely, the subcones w.r.t. the indices $j_{k,l},k\neq i$, when being currently in a cone of type i), but the projection π (in particular due to the second branch in its definition in (5.5)) folds the process down to the same hidden Markov chain $(\mathbf{Y}_k)_{k\in\mathbb{N}}$ in terms of probability. With Propositions 5.6 and 8.1 we immediately obtain:

Corollary 8.2. For almost every realisation $\big((s^{(1)},t^{(1)}),(s^{(2)},t^{(2)}),\dots\big)\in\mathcal{W}_\pi^{\mathbb{N}}$,

$$H(\mathbf{Y})=\lim_{n\to\infty}-\frac{1}{n}\log\mathbb{P}\big[\mathbf{Z}_1=(s^{(1)},t^{(1)}),\dots,\mathbf{Z}_n=(s^{(n)},t^{(n)})\big].\qquad\Box$$

The important difference between the underlying Markov chains $\big((\mathbf{t}_k,\mathbf{x}_k),(\mathbf{t}_{k+1},\mathbf{x}_{k+1})\big)_{k\in\mathbb{N}}$ and $\big((\mathbf{i}_k,\mathbf{W}_k),(\mathbf{i}_{k+1},\mathbf{W}_{k+1})\big)_{k\in\mathbb{N}}$ is that the transition matrix $\widehat{Q}_2$ has strictly positive entries, while this must not necessarily hold for the transition matrix of the Markov chain $\big((\mathbf{i}_k,\mathbf{W}_k),(\mathbf{i}_{k+1},\mathbf{W}_{k+1})\big)_{k\in\mathbb{N}}$. This property will be important later.

8.2 Proof of theorem 2.6

The crucial point will be the following lemma:

Lemma 8.3. The transition probabilities $q(w_1,w_2)$, $w_1,w_2\in\mathcal{W}_0$, vary real-analytically w.r.t. $\underline{p}\in\mathcal{P}(\underline{p}_0)$.

Proof. In order to show that $q(w_1, w_2)$ varies real-analytically in $\underline{p}$ it suffices to show analyticity of $H(ab, c)$, $ab \in \mathcal{A}^2$, $c \in \mathcal{A}$, and $\bar{L}(ab, cde)$, $d, e \in \mathcal{A}$, due to Proposition 5.1. The function $z \mapsto H(ab, c|z)$ has radius of convergence bigger than 1, which can be easily deduced from Lemma 6.1. Thus, for $\delta > 0$ small enough, we have

$$\infty > H(ab, c|1+\delta) = \sum_{n \geq 1} \mathbb{P}_{ab}[X_n = c, \forall m < n : |X_m| \geq 2](1+\delta)^n.$$

The probability $\mathbb{P}_{ab}[X_n = c, \forall m < n : |X_m| \geq 2]$ can be rewritten as

$$\sum_{\substack{n_1,\dots,n_d \geq 1: \\ n_1+\dots+n_d = n}} c(n_1, \dots, n_d) p_1^{n_1} \cdot \ldots \cdot p_d^{n_d}, \quad c(n_1, \dots, n_d) \in \mathbb{N}_0,$$

where $p_1, \dots, p_d$ correspond to the non-zero entries of the vector $\underline{p}$. Therefore,

$$H(ab, c|1+\delta) = \sum_{n \geq 1} \sum_{\substack{n_1,\dots,n_d \geq 1: \\ n_1+\dots+n_d = n}} c(n_1, \dots, n_d) (p_1(1+\delta))^{n_1} \cdot \ldots \cdot (p_d(1+\delta))^{n_d} < \infty.$$

Hence, $\underline{p}$ lies in the interior of the domain of convergence of $H(ab, c|1)$ when considered as a multivariate power series in the variables of $\mathrm{supp}(\underline{p}) = \{p_1, \dots, p_d\}$. This yields real-analyticity of $H(ab, c|1)$ in $\underline{p}$. Analyticity of $\xi(ab)$ follows now directly from its definition. One can show completely analogously that the functions $\bar{L}(ab, cde|1)$ vary also real-analytically in $\underline{p}$ since $\bar{L}(ab, cde|z)$ has also radius of convergence bigger than 1, which can also be easily deduced from Lemma 6.1. This proves the statement of the lemma. □

Now we can prove:

Proof of Theorem 2.6. The claim follows now via the equation $h = \ell \cdot H(\mathbf{Y})/\lambda$. By Lemma 8.3, the invariant probability measure ν_0 of the process $(\mathbf{W}_k)_{k \in \mathbb{N}}$ varies real-analytically in some neighbourhood of $\underline{p}_0$, since ν_0 is the solution of a linear system of equations in terms of $q(\cdot, \cdot)$; hence, λ (given in (5.8)) varies analytically.

Moreover, the transition matrix $\widehat{Q}_2$ of the process $\big((\mathbf{t}_k, \mathbf{x}_k), (\mathbf{t}_{k+1}, \mathbf{x}_{k+1})\big)_{k \in \mathbb{N}}$ has strictly positive entries. Therefore, we can apply the analyticity result for entropies of hidden Markov chains of Han and Marcus [12, Theorem 1.1] on $(\mathbf{Z}_k)_{k \in \mathbb{N}}$ and obtain together with Corollary 8.2 that $H(\mathbf{Y})$ is also real-analytic in some neighbourhood of $\underline{p}_0$; at this point it is crucial that $\widehat{Q}_2$ has strictly positive entries in order to be able to apply [12, Theorem 1.1], which was our motivation for the definition of the process $(\mathbf{t}_k, \mathbf{x}_k)_{k \in \mathbb{N}}$ and $(\mathbf{Z}_k)_{k \in \mathbb{N}}$.

Real-analyticity of ℓ can be shown completely analogously to the proof of Lemma 8.3 with the help of the formula for ℓ given in [7, Theorem 2.4]. This finishes the proof. □

A Remarks on Assumptions 2.1 and 2.4

A.1 Generalization of suffix-irreducibility

In this section we make a discussion on Assumption 2.4, where we show how to relax this condition in some way and that it cannot be dropped completely. First, recall that suffix-irreducibility leads to the fact that the process $(\mathbf{W}_k)_{k \in \mathbb{N}}$ is irreducible. One can weaken the asssumption of suffix-irreducibility to the assumption that

$$\mathbb{P}[\forall n \in \mathbb{N} : |X_n| \geq |w| \mid X_0 = w] > 0 \quad \forall w \in \mathcal{L}, \tag{A.1}$$

or equivalently that $H(ab, c|1) < 1$ for all $a, b, c \in \mathcal{A}$. This means that, for every $w \in \mathcal{L}$, there is some $ab \in \mathcal{A}^2$ such that

$$\mathbb{P}[\exists n \in \mathbb{N} : [X_n] = ab, \forall k \leq n : |X_k| \geq |w| \mid X_0 = w] > 0 \text{ and } H(ab, \cdot|1) < 1.$$

In this case the process $(\mathbf{W}_k)_{k\in\mathbb{N}}$ is not necessarily irreducible any more, but it still has a finite state space. Let $C_1, \ldots, C_r$ be the essential classes of the state space of $(\mathbf{W}_k)_{k\in\mathbb{N}}$. Then $(\mathbf{W}_k)_{k\in\mathbb{N}}$ will almost surely take only values in one of these classes up to finitely many exemptions for small $k \in \mathbb{N}$; the class depends then on the concrete realization. If we condition on the fact that $(\mathbf{W}_k)_{k\in\mathbb{N}}$ will finally enter the class C_i, then – on this event – the entropy rate h_i and the drift ℓ_i can be calculated as shown in the irreducible case and as in [7]: we just have to replace $(\mathbf{W}_k)_{k\in\mathbb{N}}$ by $(\mathbf{W}_{T+k})_{k\in\mathbb{N}}$, where T is the smallest index with $\tau(\mathbf{W}_T) \in C_i$. The overall entropy rate and drift are then given by

$$\begin{aligned} h &= \lim_{n\to\infty} \frac{1}{n}\mathbb{E}[-\log \pi(X_n)] = \sum_{i=1}^{r} h_i \cdot \mathbb{P}[(\mathbf{W}_k)_{k\in\mathbb{N}} \text{ finally enters } C_i], \\ \ell &= \lim_{n\to\infty} \frac{1}{n}\mathbb{E}[|X_n|] = \sum_{i=1}^{r} \ell_i \cdot \mathbb{P}[(\mathbf{W}_k)_{k\in\mathbb{N}} \text{ finally enters } C_i]. \end{aligned}$$

Since the probabilities $\mathbb{P}[(\mathbf{W}_k)_{k\in\mathbb{N}}$ finally enters $C_i]$ are the solutions of a finite system of linear equations with coefficients $q(\cdot,\cdot)$, they vary also analytically. Hence, condition (A.1) also implies our result on analyticity of the entropy.

If the property (A.1) does not hold, then the random walk may take some long deviations between the last entry times $\mathbf{e}_{k-1}$ and $\mathbf{e}_k$ such that $\mathbb{E}[\mathbf{e}_k - \mathbf{e}_{k-1}] = \infty$; see Example A.1 below. One can show that, in the case of infinite expectation, this leads to $\lim_{n\to\infty} k/\mathbf{e}_k = 0$, implying $\liminf_{n\to\infty} |X_n|/n = 0$; an analogous statement is shown in [9], where the proof can be adapted easily to the present context. This allows no conclusion on the entropy with our techniques, since $l(X_n) = -\log L(o, X_n|1)$ can not be compared with $-\log \pi_n(X_n)$ any more as it was done in the proof of Proposition 5.10. But we underline that this setting with deviations of expected infinite length constitutes a degenerate case.

Example A.1. Let be $\mathcal{A} = \{a, b, c, d\}$ and set

$$\begin{aligned} &p(o, a_1) = p(a_1, o) = \frac{1}{4}\ \forall a_1 \in \{a, b, c\},\ p(o, d) = \frac{1}{4},\ p(d, o) = \frac{1}{2}, \\ &p(a_1, a_1a_2) = p(a_1a_3, a_1) = \frac{1}{4}\ \forall a_1 \in \{a, b, c\}, a_2 \in \mathcal{A} \setminus \{a_1\}, a_3 \in \mathcal{A} \setminus \{a_1, d\}, \\ &p(a_1a_2, a_1a_2a_3) = \frac{1}{4}\ \forall a_1, a_2 \in \{a, b, c\}, a_1 \neq a_2, \forall a_3 \in \mathcal{A} \setminus \{a_2\}, \\ &p(ad, add) = p(bd, bdd) = p(cd, cdd) = \frac{1}{2}, \\ &p(d, dd) = p(dd, ddd) = p(dd, d) = \frac{1}{2},\ p(ad, a) = p(bd, b) = p(cd, c) = \frac{1}{2}. \end{aligned}$$

The associated graph $\mathcal{G}$ can be identified as follows: the vertex set is given by $\mathbb{T}_3 \times \mathbb{N}_0$, where $\mathbb{T}_3 = (\mathbb{Z}/2\mathbb{Z}) * (\mathbb{Z}/2\mathbb{Z}) * (\mathbb{Z}/2\mathbb{Z}) = \langle a, b, c \mid a^2 = b^2 = c^2 = 1\rangle$, and the adjacency relation is defined via $(a_1 \ldots a_k, m) \sim (b_1 \ldots b_l, n)$ if and only if

$$\begin{cases} a_1 \ldots a_k = b_1 \ldots b_l \wedge |m - n| = 1 & \text{or} \\ m = n = 0 \wedge k = l + 1 \wedge a_1 \ldots a_{k-1} = b_1 \ldots b_l \wedge a_k \neq a_{k-1} & \text{or} \\ m = n = 0 \wedge k + 1 = l \wedge a_1 \ldots a_k = b_1 \ldots b_{l-1} \wedge b_l \neq b_{l-1}. \end{cases}$$

The graph $\mathcal{G}$ can be visualized as follows: take a homogeneous tree of degree 3, where the vertices are described by words over $\{a, b, c\}$ such that two consecutive letters are different; attach to each vertex a half-line $\mathbb{N}$, where the steps on the half-line are made with equal probability of $\frac{1}{2}$; the vertices $(w,0)$ correspond to the vertices of the tree and one chooses with equal probability of $\frac{1}{4}$ one of the four neighbour vertices for the next step. This implies that the random walk will stay only for some finite time in each half-line before making a step in the tree part of $\mathcal{G}$. Moreover, it is not hard to see that the random walk converges to some infinite word over the subalphabet $\{a,b,c\}$. But it is well-known that the random walk needs in expectation infinite time to leave one of the halflines, that is, the expected time for reaching "a" when starting at "ad" is infinite. This implies that $\mathbb{E}[\mathbf{e}_k - \mathbf{e}_{k-1}] = \infty$.

A.2 Weak symmetry assumption

The purpose for introducing the weak symmetry assumption is that the random walk becomes irreducible and that the cones become strongly connected subgraphs. A weaker but still sufficient condition is given as follows: if $w_0 \in \mathcal{L}$ and $w_1, w_2 \in C(w_0)$ with

$$\mathbb{P}[\exists n \in \mathbb{N} : X_n = w_2, \forall m \leq n : X_m \in C(w_0) \mid X_0 = w_1] > 0$$

then $\mathbb{P}[\exists n \in \mathbb{N} : X_n = w_1, \forall m \leq n : X_m \in C(w_0) \mid X_0 = w_2] > 0$. Under this weaker condition the random walk still remains irreducible and the Green function's radius of convergence R is strictly bigger than 1. Also, the cones remain strongly connected and $C(w) = C(w')$ if $w' \in \partial C(w)$.

If this connectedness of cones is not satisfied then the definition of cones and coverings of cones by subcones gets more complicated. In that case the coverings depend on the boundary point from which one constructs the covering yielding coverings by possibly non-disjoint subcones. In particular, Lemma 4.2 does not necessarily hold. This would lead to a more detailed and complicated case distinction in order to get coverings by disjoint subcones. Since there will be no additional gain and the involving techniques remain the same we used weak symmetry for ease of presentation.

B Switching from the K-dependent case to the blocked letter language

In this section we make a discussion on the transition from the K-dependent case (that is, the transition probabilities depend on the last K letters and between two steps of the random walk only the last K letters may be replaced by a word of length of at most $2K$) to the blocked letter language (that is, blocking words of length of at most K to new single letters such that we are in the situation defined via (2.1)). In the *K-dependent case* the general transition probabilities have the form

$$\mathbb{P}\big[X_{n+1} = wy \mid X_n = wx\big] = p(x,y), \tag{B.1}$$

where $w, x, y \in \mathcal{A}^*$ with x being a word consisting of K letters and y being a word consisting of at most $2K$ letters.

Obviously, if the K-dependent random walk is weakly symmetric then the random walk on the blocked letter language is weakly symmetric, too. Suffix-irreducibility in the K-dependent case means that, for all $w \in \mathcal{L}$ and every $w_0 \in \mathcal{A}^K$, the random walk starting at w has positive probability to visit some word ending with w_0 by only passing through words in $\mathcal{A}_{\geq |w|}$. However, suffix-irreducibility in the K-dependent case does, in general, not necessarily yield suffix-irreducibility of the blocked letter language. But as already explained in Appendix A.1 suffix irreducibility can be relaxed by the assumption (A.1), and the blocked letter language inherits this assumption from the K-dependent case.

Finally, we want to discuss the cases when the K-dependent random walk is expanding or not. Define cones in the K-dependent case as at the beginning of Subsection 4.1.

For any $w \in \mathcal{A}^*$, denote by $[w]_K$ the last K letters. Two cones $C(w_1)$ and $C(w_2)$, $w_1, w_2 \in \mathcal{A}^*$ are then isomorphic if $C([w_1]_K) = C([w_2]_K)$. The same properties of cones and coverings (that is, nestedness or disjointness of cones, construction of coverings of cones by subcones, etc.) from Section 4 can be transferred to the K-dependent case analogously. If the graph $\mathcal{G}$ is *not* expanding in the K-dependent case then one can show analogously as in Subsection 4.2.2 that the random walk converges to one out of finitely many deterministic infinite words. In the following we will show that blocked letter language random walk is expanding if $\mathcal{G}$ is expanding in the K-dependent case. Recall that X_∞ is the infinite limiting random word of our K-dependent random walk.

Lemma B.1. If the K-dependent random walk is expanding then the support of X_∞ is infinite.

Proof. Assume that X_∞ has finite support. Choose $N \in \mathbb{N}$ large enough such that each connected component of $\mathcal{G} \setminus \{w \in \mathcal{L} \mid |w| < N\}$ (that is, remove from $\mathcal{G}$ all vertices $w \in \mathcal{L}$ with $|w| < N$ and their adjacent edges) contains in its closure only one point of the support of X_∞. Take any of these connected components and denote it by C, and take any $w_0 \in C$ with $|w_0| = N$. Then $\mathbb{P}[\forall n \geq 1 : |X_n| \geq |w_0| \mid X_0 = w_0] > 0$. Since each cone contains at least two proper subcones, we can find disjoint subcones $C(w_1), C(w_2)$ of $C(w_0)$ such that $w_1, w_2 \in \mathcal{L}$ with $|w_1| = |w_2| > |w_0| + K$. Due to condition (A.1) we have $\mathbb{P}[\forall n \geq 1 : |X_n| \geq |w_i| \mid X_0 = w_i] > 0$ for each $i \in \{1, 2\}$. We remark that this follows also from suffix-irreducibility. That is, if the random walk escapes to infinity inside $C(w_0)$ then it can escape to infinity via the cone $C(w_1)$ or via the cone $C(w_2)$, which is disjoint from $C(w_1)$. Thus, we have found two different boundary points of X_∞, which lie in the closure of C, a contradiction to our choice of N and C. Consequently, the support of X_∞ cannot be finite. □

Now we get:

Corollary B.2. If the K-dependent random walk is expanding then the associated blocked letter language random walk is also expanding.

Proof. Assume that the blocked letter language random walk is *not* expanding. Denote by $X_\infty^{(B)}$ the infinite limiting word w.r.t. the blocked letter language. Then $X_\infty^{(B)}$ is quasi-deterministic, that is, its support is a finite subset of $\mathcal{A}_B^{\mathbb{N}}$, where $\mathcal{A}_B$ is the blocked letter language alphabet. But this yields that X_∞ has also finite support in $\mathcal{A}^{\mathbb{N}}$, and this in turn implies by the previous lemma that the K-dependent case cannot be expanding. □

Hence, concerning the property "expanding" we have shown that there is no gain or loss when switching from K-dependent random walks to the blocked letter language random walk.

C Proofs

In this section we give the missing proofs of some lemmas and propositions, which we omitted earlier for sake of better readability.

Proof of Lemma 4.1.
Let be $w_1 = a_1 \dots a_m$, $w_2 = b_1 \dots b_n \in \mathcal{A}^*_{\geq 2}$ with $a_1, \dots, a_m, b_1, \dots, b_n \in \mathcal{A}$ such that $C(w_1)$ and $C(w_2)$ are isomorphic.

Proof of (1): since $C(w_1)$ and $C(w_2)$ are isomorphic we have $C([w_1]) = C([w_2])$, and thus $[w_1] = a_{m-1}a_m \in C([w_1]) = C([w_2])$. Hence, there is a path $\langle [w_2], u_1, \dots, u_k, a_{m-1}a_m \rangle$ through words $u_1, \dots, u_k \in \mathcal{A}^*_{\geq 2}$. If $w' = a_1 \dots a_{m-2}\overline{w} \in C(w_1)$ with $\overline{w} \in \mathcal{A}^*_{\geq 2}$ then there is a path $\langle w_1, w_1', \dots, w_l', w' \rangle$ through words $w_1', \dots, w_l' \in \mathcal{A}^*_{\geq |w_1|}$. This yields that w_i' has the form $w_i' = a_1 \dots a_{m-2} w_i''$ with some $w_i'' \in \mathcal{A}^*_{\geq 2}$, that is, the path $\langle a_{m-1}a_m, w_1'', \dots, w_l'', \bar{w}\rangle$ has positive probability to be performed. But this implies that

$$\langle w_2 = b_1 \dots b_{n-2}[w_2], b_1 \dots b_{n-2}u_1, \dots, b_1 \dots b_{n-2}u_k, b_1 \dots b_{n-2}a_{m-1}a_m,\\ b_1 \dots b_{n-2}w_1'', \dots, b_1 \dots b_{n-2}w_l'', b_1 \dots b_{n-2}\bar{w}\rangle$$

is a path through words in $\mathcal{A}^*_{\geq |w_2|}$, that is, $b_1 \dots b_{n-2}\bar{w} \in C(w_2)$. Thus, φ is well-defined.

Since any $w \in C(w_1)$ and its image $\varphi(w)$ differ only by different (constant) prefixes the mapping φ is obviously a bijection. Moreover, if $w = a_1 \dots a_{m-2}c_1 \dots c_k \in C(w_1)$ with $c_1, \dots, c_k \in \mathcal{A}$, $k \geq 2$, and $\hat{w} = a_1 \dots a_{m-2}c_1 \dots c_{k-2}w' \in C(w_1)$ with $w' \in \mathcal{A}^*$, $1 \leq |w'| \leq 3$, and $(k-2) + |w'| \geq 2$ (otherwise $\hat{w} \notin C(w_1)$), then

$$p(w, \hat{w}) = p(c_{k-1}c_k, w') = p(b_1 \dots b_{n-2}c_1 \dots c_k, b_1 \dots b_{n-2}c_1 \dots c_{k-2}w') = p\big(\varphi(w), \varphi(\hat{w})\big).$$

This yields (1).

Proof of (2): this follows directly from (1) by the bijection φ and the fact that the adjacency relation is given through positive single-step transition probabilities. Hence, $C(w_1)$ and $C(w_2)$ are isomorphic as subgraphs of $\mathcal{G}$. □

Proof of Lemma 4.2.
Let be $w_1, w_2 \in \mathcal{A}^*_{\geq 2}$. W.l.o.g. assume that $|w_1| \leq |w_2|$. Moreover, assume that the cones $C(w_1)$ and $C(w_2)$ are not nested in each other and that $C(w_1) \cap C(w_2) \neq \emptyset$. Let be $w_0 \in C(w_1) \cap C(w_2)$. Then there is a path $\langle w_1, w_1', \dots, w_k', w_0\rangle$ through words $w_1', \dots, w_k' \in \mathcal{A}^*_{\geq |w_1|}$ and there is a path $\langle w_2, w_1'', \dots, w_l'', w_0\rangle$ through words $w_1'', \dots, w_l'' \in \mathcal{A}^*_{\geq |w_2|} \subseteq \mathcal{A}^*_{\geq |w_1|}$. By weak symmetry, there is a path $\langle w_1, w_1', \dots, w_k', w_0, w_l'', \dots, w_1'', w_2\rangle$ through words in $\mathcal{A}^*_{\geq |w_1|}$, and hence $w_2 \in C(w_1)$ which in turn implies $C(w_2) \subseteq C(w_1)$, a contradiction. This yields the first part of the lemma.

In order to prove the second part assume that $|w_1| = |w_2|$ and w.l.o.g. $C(w_1) \subseteq C(w_2)$. It remains to show that we have then $C(w_1) = C(w_2)$. Since $w_1 \in C(w_2)$ there is a path $\langle w_2, \bar{w}_1, \dots, \bar{w}_m, w_1\rangle$ through words $\bar{w}_1, \dots, \bar{w}_m \in \mathcal{A}^*_{\geq |w_2|}$. If $w \in C(w_2)$ then there is a path $\langle w_2, \hat{w}_1, \dots, \hat{w}_n, w\rangle$ through words $\hat{w}_1, \dots, \hat{w}_n \in \mathcal{A}^*_{\geq |w_2|}$. Thus, there is a path

$$\langle w_1, \bar{w}_m, \dots, \bar{w}_1, w_2, \hat{w}_1, \dots, \hat{w}_n, w\rangle$$

though words in $\mathcal{A}^*_{\geq |w_2|} = \mathcal{A}^*_{\geq |w_1|}$. Hence, $C(w_2) \subseteq C(w_1)$ which yields $C(w_2) = C(w_1)$. □

For the next proof we need the following properties: if $a_1b_1, a_2b_2 \in \mathcal{A}^2$ satisfy $\tau(C(a_1b_1)) = \tau(C(a_2b_2))$ then we have $C(a_1b_1) = C(a_2b_2)$ (see Lemma 4.2) and therefore $a_2b_2 \in C(a_1b_1)$. In this case we also have $\mathbb{L}(a_1b_1, w) > 0$ for $w \in \mathcal{A}^*_{\geq 3}$ if and only if $\mathbb{L}(a_2b_2, w) > 0$. This follows from the simple fact that $a_2b_2 \in C(a_1b_1)$ implies that there are paths from a_1b_1 to a_2b_2 (and vice versa) through words in $\mathcal{A}^*_{\geq 2}$.

Proof of Lemma 5.2.
By definition, we obviously have $\mathrm{supp}(\mathbb{P}[\mathbf{W}_1 = \cdot]) = \mathcal{W}_0$. For $k > 1$ we show both inclusions. Let be $y \in \mathcal{W}_0$. Then there are $w_0 \in \mathcal{A}^*$ and $ab \in \mathcal{A}^2$ with $w_0ab \in \bigcup_{j=1}^{n_0} \partial C_j^{(0)}$ and $w_0y \in \mathcal{S}(w_0ab)$ and

$$\begin{aligned}\mathbb{P}[\mathbf{W}_0 = w_0ab, \mathbf{W}_1 = y] &= \sum_{w' \in \mathcal{L} \setminus C(w_0ab)} G(o, w') \cdot p(w', w_0ab) \cdot \mathbb{L}(w_0ab, w_0y) \cdot \xi([y]) \\ &= \sum_{w' \in \mathcal{L} \setminus C(w_0ab)} G(o, w') \cdot p(w', w_0ab) \cdot \mathbb{L}(ab, y) \cdot \xi([y]) > 0.\end{aligned}$$

Take now any $\bar{w}\bar{a}\bar{b} \in \mathrm{supp}(\mathbb{P}[X_{\mathbf{e}_{k-2}} = \cdot])$. Since the covering of every cone contains subcones of all different types, the cone $C(\bar{w}\bar{a}\bar{b})$ has in its covering a cone of type $\tau(C(ab))$. Hence, there are $w_k \in \mathcal{A}^*$, $a_k b_k \in \mathcal{A}^2$ with $\bar{w} w_k a_k b_k \in \mathcal{S}(\bar{w}\bar{a}\bar{b})$, $\tau(C(a_k b_k)) = \tau(C(ab))$ and $m_k \in \mathbb{N}$ such that $p^{(m_k)}(o, \bar{w} w_k a_k b_k) > 0$. Thus,

$$\begin{aligned}
\mathbb{P}[\mathbf{W}_k = y] &\geq \mathbb{P}[X_{\mathbf{e}_{k-1}} = \bar{w} w_k a_k b_k, \mathbf{W}_k = y] \\
&= \sum_{w' \in \mathcal{L} \setminus C(\bar{w} w_k a_k b_k)} G(o, w') \cdot p(w', \bar{w} w_k a_k b_k) \cdot \mathbb{L}(\bar{w} w_k a_k b_k, \bar{w} w_k y) \cdot \xi([y]) \\
&= \sum_{w' \in \mathcal{L} \setminus C(\bar{w} w_k a_k b_k)} G(o, w') \cdot p(w', \bar{w} w_k a_k b_k) \cdot \mathbb{L}(a_k b_k, y) \cdot \xi([y]).
\end{aligned}$$

By the remark before the lemma, we have $\mathbb{L}(a_k b_k, y) > 0$ and therefore $\mathbb{P}[\mathbf{W}_k = y] > 0$, yielding $\mathcal{W}_0 \subseteq \mathrm{supp}(\mathbb{P}[\mathbf{W}_k = \cdot])$.

For the other direction, take any $y \in \mathrm{supp}(\mathbb{P}[\mathbf{W}_k = \cdot])$. Then there is some $w_{k-1}ab \in \mathcal{L}$ such that

$$\begin{aligned}
0 &< \mathbb{P}[X_{\mathbf{e}_{k-1}} = w_{k-1}ab, X_{\mathbf{e}_k} = w_{k-1}y] \\
&= \sum_{w' \in \mathcal{L} \setminus C(w_{k-1}ab)} G(o, w') \cdot p(w', w_{k-1}ab) \cdot \mathbb{L}(w_{k-1}ab, w_{k-1}y) \cdot \xi([y]).
\end{aligned}$$

In particular, $\mathbb{L}(ab, y) > 0$. Since the initial covering of $\mathcal{L}$ contains a cone of type $\tau(C(ab))$ there are $w_0 \in \mathcal{A}^*$, $a_0 b_0 \in \mathcal{A}^2$ and some $m \in \mathbb{N}$ such that $w_0 a_0 b_0 \in \bigcup_{i=1}^{n_0} \partial C_i^{(0)}$, $\tau(C(a_0 b_0)) = \tau(C(ab))$ and $p^{(m)}(o, w_0 a_0 b_0) > 0$. Observe again that $\mathbb{L}(a_0 b_0, y) > 0$ by the remark before the lemma. Therefore,

$$\begin{aligned}
\mathbb{P}[\mathbf{W}_1 = y] &\geq \mathbb{P}[\mathbf{W}_0 = w_0 a_0 b_0, \mathbf{W}_1 = y] = \mathbb{P}[X_{\mathbf{e}_0} = w_0 a_0 b_0, \mathbf{W}_1 = y] \\
&= \sum_{w' \in \mathcal{L} \setminus C(w_0 a_0 b_0)} G(o, w') \cdot p(w', w_0 a_0 b_0) \cdot \mathbb{L}(w_0 a_0 b_0, w_0 y) \cdot \xi([y]) \\
&= \sum_{w' \in \mathcal{L} \setminus C(w_0 a_0 b_0)} G(o, w') \cdot p(w', w_0 a_0 b_0) \cdot \mathbb{L}(a_0 b_0, y) \cdot \xi([y]) > 0.
\end{aligned}$$

This yields $\mathrm{supp}(\mathbb{P}[\mathbf{W}_k = \cdot]) \subseteq \mathrm{supp}(\mathbb{P}[\mathbf{W}_1 = \cdot]) = \mathcal{W}_0$ and the claim of the lemma follows. □

Proof of Proposition 5.4.
It remains to show that the support of each $(\mathbf{i}_k, \mathbf{W}_k)$ equals $\mathcal{W}$ and that $\big((\mathbf{i}_k, \mathbf{W}_k)\big)_{k \in \mathbb{N}}$ is positive recurrent and aperiodic.

First, we show that $\mathrm{supp}(\mathbb{P}[(\mathbf{i}_k, \mathbf{W}_k) = \cdot]) = \mathcal{W}$ for $k \geq 1$. For this purpose, let be $(j_{i,n}, x) \in \mathrm{supp}(\mathbb{P}[(\mathbf{i}_k, \mathbf{W}_k) = \cdot])$. Then there is some $w_{k-1} a_{k-1} b_{k-1} \in \mathcal{L}$ with

$$\begin{aligned}
&\mathbb{P}[X_{\mathbf{e}_{k-1}} = w_{k-1} a_{k-1} b_{k-1}, \mathbf{W}_k = x] \\
= &\sum_{w' \in \mathcal{L} \setminus C(w_{k-1} a_{k-1} b_{k-1})} G(o, w') p(w', w_{k-1} a_{k-1} b_{k-1}) \mathbb{L}(a_{k-1} b_{k-1}, x) \xi([x]) > 0,
\end{aligned}$$

$\tau(C(a_{k-1} b_{k-1})) = i$ and $C(x)$ being the n-th subcone of type j in the covering of the cone $C(a_{k-1} b_{k-1})$. If $k = 1$ then $(j_{i,n}, x) \in \mathcal{W}$. In the case $k > 1$ take any $w_0 a_0 b_0 \in \mathcal{L}$ with $\mathbb{P}[\mathbf{W}_0 = w_0 a_0 b_0] > 0$ and $\tau(C(w_0 a_0 b_0)) = i$. Since $a_{k-1} b_{k-1} \in C(a_0 b_0)$ we also have $\mathbb{L}(a_0 b_0, x) > 0$ since $\mathbb{L}(a_{k-1} b_{k-1}, x) > 0$ (recall the remark before Lemma 5.2). Then:

$$\mathbb{P}[\mathbf{W}_0 = w_0 a_0 b_0, \mathbf{W}_1 = x] = \sum_{w' \in \mathcal{L} \setminus C(w_0 a_0 b_0)} G(o, w') p(w', w_0 a_0 b_0) \mathbb{L}(a_0 b_0, x) \xi([x]) > 0,$$

yielding $(j_{i,n}, x) \in \mathcal{W}$.

For the other inclusion, let be $(j_{i,n}, x) \in \mathcal{W}$. Then there is some $w_0a_0b_0 \in \mathcal{L}$ with

$$\mathbb{P}[\mathbf{W}_0 = w_0a_0b_0, \mathbf{W}_1 = x] = \sum_{w' \in \mathcal{L}\setminus C(w_0a_0b_0)} G(o,w')p(w', w_0a_0b_0)\mathbb{L}(a_0b_0, x)\xi([x]) > 0,$$

$\tau(C(a_0b_0)) = i$ and $C(x)$ being the n-th subcone of type j in the covering of $C(a_0b_0)$. If $k = 1$ then $(j_{i,n}, x) \in \operatorname{supp}(\mathbb{P}[(\mathbf{i}_1, \mathbf{W}_1) = \cdot])$. In the case $k > 1$ take any $w_{k-2}a_{k-2}b_{k-2} \in \mathcal{L}$ with $\mathbb{P}[X_{\mathbf{e}_{k-2}} = w_{k-2}a_{k-2}b_{k-2}] > 0$. Then $C(w_{k-2}a_{k-2}b_{k-2})$ has in its covering a subcone $C(w_{k-1}a_{k-1}b_{k-1})$ of type i. Since $a_{k-1}b_{k-1} \in C(a_0b_0)$ we have $\mathbb{L}(a_{k-1}b_{k-1}, x) > 0$ due to $\mathbb{L}(a_0b_0, x) > 0$ (once again recall the remark before Lemma 5.2) and $C(x)$ is the n-th subcone of type j in the covering of $C(a_{k-1}b_{k-1}) = C(a_0b_0)$. Hence,

$$\begin{aligned}
&\mathbb{P}[(\mathbf{i}_k, \mathbf{W}_k) = (j_{i,n}, x)] \geq \mathbb{P}[X_{\mathbf{e}_{k-1}} = w_{k-1}a_{k-1}b_{k-1}, X_{\mathbf{e}_k} = w_{k-1}x] \\
\geq &\sum_{w' \in \mathcal{L}\setminus C(w_{k-1}a_{k-1}b_{k-1})} G(o,w')p(w', w_{k-1}a_{k-1}b_{k-1})\mathbb{L}(a_{k-1}b_{k-1}, x)\xi([x]) > 0,
\end{aligned}$$

yielding $\mathcal{W} \subseteq \operatorname{supp}(\mathbb{P}[(\mathbf{i}_k, \mathbf{W}_k) = \cdot])$, and therefore $\mathcal{W} = \operatorname{supp}(\mathbb{P}[(\mathbf{i}_k, \mathbf{W}_k) = \cdot])$.

The next task is to show irreducibility, which implies positive recurrence due to finiteness of $\mathcal{W}$. Let be $(i_{m,n}, w_1)$, $(j_{s,t}, w_2) \in \mathcal{W}$. Take any $\bar{w} \in \mathcal{W}_0$ such that $q(w_1, \bar{w}) > 0$ and $\tau(C(\bar{w})) = s$, which exists by construction of coverings. Then $w_2 \in \partial C_{j_{s,t}}([\bar{w}])$, that is, $C(w_2)$ is the t-th subcone of type j in the covering of $C([\bar{w}])$, yielding $q(\bar{w}, w_2) > 0$. Hence,

$$\begin{aligned}
&\mathbb{P}[(\mathbf{i}_3, \mathbf{W}_3) = (j_{s,t}, w_2) \mid (\mathbf{i}_1, \mathbf{W}_1) = (i_{m,n}, w_1)] \\
\geq\ &\mathbb{P}[\mathbf{W}_3 = w_2, \mathbf{W}_2 = \bar{w} \mid (\mathbf{i}_1, \mathbf{W}_1) = (i_{m,n}, w_1)] \\
=\ &q(w_1, \bar{w}) \cdot q(\bar{w}, w_2) > 0.
\end{aligned} \tag{C.1}$$

Here, we used the fact that $\mathbf{i}_3 = j_{s,t}$ is uniquely determined by $w_1, \bar{w}, w_2$ and that this probability does not depend on m and n. This yields irreducbility of the process $\big((\mathbf{i}_k, \mathbf{W}_k)\big)_{k\in\mathbb{N}}$.

It follows that the period of the process is at most 2. In order to see aperiodicity, take any $\underline{w_1}, \underline{w_2} \in \mathcal{W}$ with $\mathbb{P}[(\mathbf{i}_2, \mathbf{W}_2) = \underline{w_2} \mid (\mathbf{i}_1, \mathbf{W}_1) = \underline{w_1}] > 0$. Then we get analogously to (C.1):

$$\begin{aligned}
&\mathbb{P}[(\mathbf{i}_4, \mathbf{W}_4) = \underline{w_1}, (\mathbf{i}_2, \mathbf{W}_2) = \underline{w_2} \mid (\mathbf{i}_1, \mathbf{W}_1) = \underline{w_1}] \\
=\ &\mathbb{P}[(\mathbf{i}_2, \mathbf{W}_2) = \underline{w_2} \mid (\mathbf{i}_1, \mathbf{W}_1) = \underline{w_1}] \cdot \mathbb{P}[(\mathbf{i}_4, \mathbf{W}_4) = \underline{w_1} \mid (\mathbf{i}_2, \mathbf{W}_2) = \underline{w_2}] > 0.
\end{aligned}$$

That is, the period of the process is 1. This finishes the proof. □

The following lemma was used in the proof of Proposition 5.8:

Lemma C.1. For almost every realisation $(x_1, \underline{y}_1, \underline{y}_2, \dots)$ of $(X_{\mathbf{e}_1}, \mathbf{Y}_1, \mathbf{Y}_2, \dots)$,

$$H(\mathbf{Y}) = \lim_{k\to\infty} -\frac{1}{k} \log \mathbb{P}\big[C(X_{\mathbf{e}_1}) = C(x_1), \mathbf{Y}_1 = \underline{y}_1, \dots, \mathbf{Y}_k = \underline{y}_k\big].$$

Proof. We recall the notation from the proof of Proposition 5.8: let be $k \in \mathbb{N}$ and assume for the moment that $\mathbf{W}_l = y_la_lb_l$, where $y_l \in \mathcal{A}^* \setminus \{o\}$ and $a_lb_l \in \mathcal{A}^2$ for $0 \leq l \leq k$. That is, $X_{\mathbf{e}_l} = y_0y_1 \dots y_la_lb_l$. We write $\mathbf{Y}_1 = (j, t^{(1)})$, where $j = \tau(C(a_1b_1))$, and $\mathbf{Y}_l = (s^{(l)}, t^{(l)})$ for $2 \leq l \leq k$, where the values of $s^{(2)}, \dots, s^{(k-1)}$ and $t^{(1)}, \dots, t^{(k-1)}$ are determined by the values of $\mathbf{W}_l = y_la_lb_l$. Vice versa, given $X_{\mathbf{e}_1}$ the values of $s^{(2)}, \dots, s^{(k-1)}$ and $t^{(1)}, \dots, t^{(k-1)}$ determine uniquely the cones $C(y_la_lb_l)$: indeed, $X_{\mathbf{e}_1}$ and $t^{(1)}$ determine uniquely $C(X_{\mathbf{e}_2})$ and therefore also $C(\mathbf{W}_2) = C(y_2a_2b_2)$; inductively, given $C(X_{\mathbf{e}_l})$ of type $s^{(l)}$ then $t^{(l)}$ determines uniquely $C(X_{\mathbf{e}_{l+1}})$ and $C(\mathbf{W}_{l+1}) = C(y_{l+1}a_{l+1}b_{l+1})$. We mark it by $(*)$ when we make use of this "transition".

Recall that the covering of $\mathcal{L}$ consists of n_0 subcones $C_i^{(0)}$, $1 \leq i \leq n_0$. Each $C_i^{(0)}$ has again a covering consisting of $n(\tau(C_i^{(0)}), j)$ subcones of type j. We enumerate all these subcones of type j by $C_{j,k}^{(1)}$ with $1 \leq k \leq N_j := \sum_{i=1}^{n_0} n(\tau(C_i^{(0)}), j)$, that is, we enumerate all subcones of type j which appear in the coverings of all $C_i^{(0)}$, $1 \leq i \leq n_0$.

Since $\mathcal{W}_0$ is finite, there is some constant $c > 0$ such that

$$c \cdot \mathbb{P}[X_{\mathbf{e}_1} = x] \leq \mathbb{P}[X_{\mathbf{e}_1} = y]$$

for all $x, y \in \bigcup_{k=1}^{N_j} \partial C_{j,k}^{(1)} \subseteq \mathrm{supp}(\mathbb{P}[X_{\mathbf{e}_1} = \cdot])$.

In the following we will show that $\mathbb{P}\big[C(X_{\mathbf{e}_1}) = C(x_1), \mathbf{Y}_1 = \underline{y}_1, \ldots, \mathbf{Y}_k = \underline{y}_k\big]$ is comparable with $\mathbb{P}\big[\mathbf{Y}_1 = \underline{y}_1, \ldots, \mathbf{Y}_k = \underline{y}_k\big]$, which proves the claim. First, we have for $k \geq 2$:

$$\begin{aligned}
& N_j \cdot \mathbb{P}\big[X_{\mathbf{e}_1} \in C(y_0 y_1 a_1 b_1), \mathbf{Y}_1 = (j, t^{(1)}), \mathbf{Y}_2 = (s^{(2)}, t^{(2)}), \ldots, \mathbf{Y}_{k-1} = (s^{(k-1)}, t^{(k-1)})\big] \\
\stackrel{(*)}{=} \; & N_j \cdot \sum_{x \in \partial C(y_0 y_1 a_1 b_1)} \sum_{\substack{w_2, \ldots, w_k \in \mathcal{W}_0: \\ w_i \in \partial C(y_i a_i b_i) \\ \text{for all } 2 \leq i \leq k}} \mathbb{P}[X_{\mathbf{e}_1} = x, X_{\mathbf{e}_2} = y_0 y_1 w_2, \ldots, X_{\mathbf{e}_k} = y_0 \ldots y_{k-1} w_k] \\
= \; & N_j \cdot \sum_{\substack{x \in \partial C(y_0 y_1 a_1 b_1); \\ w_2, \ldots, w_k \in \mathcal{W}_0: \\ w_i \in \partial C(y_i a_i b_i) \\ \text{for all } 2 \leq i \leq k}} \mathbb{P}[X_{\mathbf{e}_1} = x] \mathbb{P}[X_{\mathbf{e}_2} = y_0 y_1 w_2, \ldots, X_{\mathbf{e}_k} = y_0 \ldots y_{k-1} w_k \mid X_{\mathbf{e}_1} = x] \\
= \; & N_j \cdot \sum_{x \in \partial C(y_0 y_1 a_1 b_1)} \sum_{\substack{w_2, \ldots, w_k \in \mathcal{W}_0: \\ w_i \in \partial C(y_i a_i b_i) \\ \text{for all } 2 \leq i \leq k}} \mathbb{P}[X_{\mathbf{e}_1} = x] q(y_1[x], w_2) \prod_{i=3}^{k} q(w_{i-1}, w_i) \\
= \; & \sum_{l=1}^{N_j} \sum_{x \in \partial C(y_0 y_1 a_1 b_1)} \sum_{\substack{w_2, \ldots, w_k \in \mathcal{W}_0: \\ w_i \in \partial C(y_i a_i b_i) \\ \text{for all } 2 \leq i \leq k}} \mathbb{P}[X_{\mathbf{e}_1} = x] \mathbb{P}\big[\mathbf{W}_2 = w_2 \big| [X_{\mathbf{e}_1}] = [x]\big] \prod_{i=3}^{k} q(w_{i-1}, w_i).
\end{aligned}$$

For a moment, let be $\partial C(y_0 y_1 a_1 b_1) = \{y_0 y_1 c_1 d_1, \ldots, y_0 y_1 c_\kappa d_\kappa\}$. Then for all $l \in \{1, \ldots, N_j\}$ there is some $v_l \in \mathcal{A}^*$ such that $\partial C_{j,l}^{(1)} = \{v_l c_1 d_1, \ldots, v_l c_\kappa d_\kappa\}$. Therefore, for every $x \in \partial C(y_0 y_1 a_1 b_1)$ and each $l \in \{1, \ldots, N_j\}$ there is exactly one $\hat{x}_l \in \partial C_{j,l}^{(1)}$ with $[\hat{x}_l] = [x]$, $\mathbb{P}[X_{\mathbf{e}_1} = x] \geq c \cdot \mathbb{P}[X_{\mathbf{e}_1} = \hat{x}_l]$ and $\mathbb{P}\big[\mathbf{W}_2 = w_2 \big| [X_{\mathbf{e}_1}] = [x]\big] = \mathbb{P}\big[\mathbf{W}_2 = w_2 \big| [X_{\mathbf{e}_1}] = [\hat{x}_l]\big]$ for all $w_2 \in \mathcal{W}_0$. The last equation follows from the fact that the probabilities depend on $X_{\mathbf{e}_1}$ only by its last two letters $[X_{\mathbf{e}_1}]$ in the condition. We write $\hat{x}_l$ for this mapping $(x, l) \mapsto \hat{x}_l$. Hence,

$$\begin{aligned}
& N_j \cdot \mathbb{P}\left[\begin{array}{c} X_{\mathbf{e}_1} \in C(y_0 y_1 a_1 b_1), \mathbf{Y}_1 = (j, t^{(1)}), \\ \mathbf{Y}_2 = (s^{(2)}, t^{(2)}), \ldots, \mathbf{Y}_{k-1} = (s^{(k-1)}, t^{(k-1)}) \end{array}\right] \\
\geq \; & \sum_{l=1}^{N_j} \sum_{x \in \partial C(y_0 y_1 a_1 b_1)} \sum_{\substack{w_2, \ldots, w_k \in \mathcal{W}_0: \\ w_i \in \partial C(y_i a_i b_i) \\ \text{for all } 2 \leq i \leq k}} c \mathbb{P}[X_{\mathbf{e}_1} = \hat{x}_l] \mathbb{P}\big[\mathbf{W}_2 = w_2 \mid [X_{\mathbf{e}_1}] = [\hat{x}_1]\big] \prod_{i=3}^{k} q(w_{i-1}, w_i) \\
= \; & \sum_{l=1}^{N_j} \sum_{w \in \partial C_{j,l}^{(1)}} \sum_{\substack{w_2, \ldots, w_k \in \mathcal{W}_0: \\ w_i \in \partial C(y_i a_i b_i) \\ \text{for all } 2 \leq i \leq k}} c \cdot \mathbb{P}[X_{\mathbf{e}_1} = w] \cdot \mathbb{P}\big[\mathbf{W}_2 = w_2 \big| [X_{\mathbf{e}_1}] = [w]\big] \cdot \prod_{i=3}^{k} q(w_{i-1}, w_i) \\
= \; & c \cdot \mathbb{P}\big[\mathbf{Y}_1 = (j, t^{(1)}), \mathbf{Y}_2 = (s^{(2)}, t^{(2)}), \ldots, \mathbf{Y}_{k-1} = (s^{(k-1)}, t^{(k-1)})\big].
\end{aligned}$$

Vice versa, we obviously have

$$\begin{aligned}
& \mathbb{P}\big[X_{\mathbf{e}_1} \in C(y_0 y_1 a_1 b_1), \mathbf{Y}_1 = (j, t^{(1)}), \mathbf{Y}_2 = (s^{(2)}, t^{(2)}), \dots, \mathbf{Y}_{k-1} = (s^{(k-1)}, t^{(k-1)})\big] \\
\leq\ & \mathbb{P}\big[\mathbf{Y}_1 = (j, t^{(1)}), \mathbf{Y}_2 = (s^{(2)}, t^{(2)}), \dots, \mathbf{Y}_{k-1} = (s^{(k-1)}, t^{(k-1)})\big].
\end{aligned}$$

This proves the claim. □

Proof of Proposition 8.1. Let be $(s^{(1)}, t^{(1)}), \dots, (s^{(n)}, t^{(n)}) \in \mathcal{W}_\pi$. We prove the claim by induction on n. First, let be $j, s \in \mathcal{I}$ and $t^{(1)} = j_m$ with $2 \leq m \leq n(s,j)$, and let $a_0 b_0, ab \in \mathcal{A}^2$ with $\tau(C(a_0 b_0)) = s$ and $\tau(C(ab)) = j$. If $C_{j,m}$ is the m-th cone of type j in the covering of $C(a_0 b_0)$ then there is a unique word $\bar{x}_0 = \bar{x}_0^{[s,j,m,ab]} \in \mathcal{A}^*$ with $\bar{x}_0 ab \in \partial C_{j,m}$. With this notation we get:

$$\begin{aligned}
\mathbb{P}\big[\mathbf{Y}_1 = (s, j_m), [\mathbf{W}_2] = ab\big] &= \sum_{(u_{k,l}, x) \in \mathcal{W}: u = s} \mathbb{P}[(\mathbf{i}_1, \mathbf{W}_1) = (s_{k,l}, x)] \cdot q(x, \bar{x}_0 ab) \\
&= \sum_{(u_{k,l}, x) \in \mathcal{W}: u = s} \hat{\mu}_1(s_{k,l}, x) \hat{q}\big((s_{k,l}, x), (j_{s,m}, \bar{x}_0 ab)\big) \\
&= \mathbb{P}\big[\mathbf{Z}_1 = (s, j_m), [\mathbf{x}_2] = ab\big].
\end{aligned}$$

Now we turn to the case $t^{(1)} = j_1$. Once again, if $C_{j,1}$ is the first cone of type j in the covering of $C(a_0 b_0)$ then there is some unique $\bar{x}_0 = \bar{x}_0^{[s,j,1,ab]} \in \mathcal{A}^*$ with $\bar{x}_0 ab \in \partial C_{j,1}$. We get:

$$\begin{aligned}
& \mathbb{P}\big[\mathbf{Z}_1 = (s, j_1), [\mathbf{x}_2] = ab\big] \\
&= \sum_{(u_{k,l}, x) \in \mathcal{W}: u = s} \hat{\mu}_1(s_{k,l}, x) \Big[\hat{q}\big((s_{k,l}, x), (j_{s,1}, \bar{x}_0 ab)\big) + \sum_{\substack{(t_{p,q}, y) \in \mathcal{W}: \\ t = j, p \neq s, [y] = ab}} \hat{q}\big((s_{k,l}, x), (j_{p,q}, y)\big)\Big] \\
&= \sum_{\substack{(u_{k,l}, x) \in \mathcal{W}: \\ u = s}} \hat{\mu}_1(s_{k,l}, x) \left[\frac{q\big((s_{k,l}, x), (j_{s,1}, \bar{x}_0 ab)\big)}{\#\{t_{\kappa_1,\kappa_2} \mid \kappa_1 \neq s, ab\} + 1} + \sum_{\substack{(t_{p,q}, y) \in \mathcal{W}: \\ t = j, p \neq s, \\ [y] = ab}} \frac{q\big((s_{k,l}, x), (j_{s,1}, \bar{x}_0 ab)\big)}{\#\{t_{\kappa_1,\kappa_2} \mid \kappa_1 \neq s, ab\} + 1}\right] \\
&= \sum_{(u_{k,l}, x) \in \mathcal{W}: u = s} \mathbb{P}\big[(\mathbf{i}_1, \mathbf{W}_1) = (s_{k,l}, x)\big] \cdot q(x, \bar{x}_0 ab) = \mathbb{P}\big[\mathbf{Y}_1 = (s, j_1), [\mathbf{W}_2] = ab\big].
\end{aligned}$$

Now, in both cases we obtain

$$\begin{aligned}
\mathbb{P}\big[\mathbf{Z}_1 = (s, t^{(1)})\big] &= \sum_{ab \in \mathcal{A}^2} \mathbb{P}\big[\mathbf{Z}_1 = (s, t^{(1)}), [\mathbf{x}_2] = ab\big] \\
&= \sum_{ab \in \mathcal{A}^2} \mathbb{P}\big[\mathbf{Y}_1 = (s, t^{(1)}), [\mathbf{W}_2] = ab\big] = \mathbb{P}\big[\mathbf{Y}_1 = (s, t^{(1)})\big].
\end{aligned}$$

We now perform the induction step where we will use the induction assumption

$$\begin{aligned}
& \mathbb{P}\big[\mathbf{Y}_1 = (s^{(1)}, t^{(1)}), \dots, \mathbf{Y}_n = (s^{(n)}, t^{(n)}), [\mathbf{W}_{n+1}] = ab\big] \\
=\ & \mathbb{P}\big[\mathbf{Z}_1 = (s^{(1)}, t^{(1)}), \dots, \mathbf{Z}_n = (s^{(n)}, t^{(n)}), [\mathbf{x}_{n+1}] = ab\big].
\end{aligned} \tag{C.2}$$

First, consider the case $(s^{(n+1)}, t^{(n+1)}) = (s, j_m)$ with $s, j \in \mathcal{I}$ and $2 \leq m \leq n(s,j)$. This implies that $\mathbf{t}_{n+1}$ has the form $s_{*,*}$ and $\mathbf{t}_{n+2} = j_{s,m}$. Let $C_{j,m}$ be the m-th cone of type j in the covering of $C(a_0 b_0)$, where $a_0 b_0 \in \mathcal{A}^2$ with $\tau(C(a_0 b_0)) = s$. If $ab \in \mathcal{A}^2$ with $\tau(C(ab)) = j$ then there is some unique $\bar{x}_0 = \bar{x}_0^{[s,j,m,ab]} \in \mathcal{A}^*$ with $\bar{x}_0 ab \in \partial C_{j,m}$. In this case we obtain:

$$\mathbb{P}\big[\mathbf{Z}_1 = (s^{(1)}, t^{(1)}), \dots, \mathbf{Z}_{n+1} = (s^{(n+1)}, j_m), [\mathbf{x}_{n+1}] = a_0 b_0, [\mathbf{x}_{n+2}] = ab\big]$$

$$\begin{aligned}
&= \sum_{\substack{(u_{k,l},w_0)\in\mathcal{W}:\\ u=s,[w_0]=a_0b_0}} \mathbb{P}\bigl[\mathbf{Z}_1=(s^{(1)},t^{(1)}),\dots,\mathbf{Z}_n=(s^{(n)},t^{(n)}),\mathbf{t}_{k+1}=u_{k,l},\mathbf{x}_{n+1}=w_0\bigr]\\
&\qquad \cdot \hat{q}\bigl((s_{k,l},w_0),(j_{s,m},\bar{x}_0ab)\bigr)\\
&= \mathbb{P}\bigl[\mathbf{Z}_1=(s^{(1)},t^{(1)}),\dots,\mathbf{Z}_n=(s^{(n)},t^{(n)}),[\mathbf{x}_{n+1}]=a_0b_0\bigr]\frac{\xi(ab)}{\xi(a_0b_0)}\mathbb{L}(a_0b_0,\bar{x}_0ab)\\
&= \mathbb{P}\bigl[\mathbf{Y}_1=(s^{(1)},t^{(1)}),\dots,\mathbf{Y}_n=(s^{(n)},t^{(n)}),[\mathbf{W}_{n+1}]=a_0b_0\bigr]\frac{\xi(ab)}{\xi(a_0b_0)}\mathbb{L}(a_0b_0,\bar{x}_0ab)\\
&= \mathbb{P}\bigl[\mathbf{Y}_1=(s^{(1)},t^{(1)}),\dots,\mathbf{Y}_{n+1}=(s^{(n+1)},j_m),[\mathbf{W}_{n+1}]=a_0b_0,[\mathbf{W}_{n+2}]=ab\bigr].
\end{aligned}$$

Now we turn to the case $(s^{(n+1)},t^{(n+1)})=(s,j_1)$. This implies again that $\mathbf{t}_{n+1}$ has the form $s_{*,*}$. Once again, if $C_{j,1}$ is the first cone of type j in the covering of $C(a_0b_0)$ (of type s) then there is some unique $\bar{x}_0=\bar{x}_0^{[s,j,1,ab]}\in\mathcal{A}^*$ with $\bar{x}_0ab\in\partial C_{j,1}$. We get by distinguishing whether $t^{(n+1)}=j_1$ arises from $\mathbf{t}_{n+2}=j_{s,1}$ or $\mathbf{t}_{n+2}=j_{k,l}$ with $k\neq s$:

$$\begin{aligned}
&\mathbb{P}\bigl[\mathbf{Z}_1=(s^{(1)},t^{(1)}),\dots,\mathbf{Z}_{n+1}=(s^{(n+1)},j_1),[\mathbf{x}_{n+1}]=a_0b_0,[\mathbf{x}_{n+2}]=ab\bigr]\\
=\ & \sum_{\substack{(u_{p,q},w_0)\in\mathcal{W}:\\ u=s,[w_0]=a_0b_0}} \mathbb{P}\bigl[\mathbf{Z}_1=(s^{(1)},t^{(1)}),\dots,\mathbf{Z}_n=(s^{(n)},t^{(n)}),\mathbf{t}_{n+1}=u_{p,q},\mathbf{x}_{n+1}=w_0\bigr]\\
&\qquad \cdot\Bigl(\hat{q}\bigl((s_{p,q},w_0),(j_{s,1},\bar{x}_0ab)\bigr)+\sum_{\substack{(t_{k,l},y)\in\mathcal{W}:\\ t=j,k\neq s,[y]=ab}}\hat{q}\bigl((s_{p,q},w_0),(j_{k,l},y)\bigr)\Bigr)\\
=\ & \mathbb{P}\bigl[\mathbf{Z}_1=(s^{(1)},t^{(1)}),\dots,\mathbf{Z}_n=(s^{(n)},t^{(n)}),[\mathbf{x}_{n+1}]=a_0b_0\bigr]\\
&\quad\cdot\Biggl[\frac{\xi(ab)}{\xi(a_0b_0)}\frac{\mathbb{L}(a_0b_0,\bar{x}_0ab)}{\#\{j_{k,l}\mid k\neq s,ab\}+1}+\sum_{\substack{(t_{k,l},y)\in\mathcal{W}:\\ t=j,k\neq s,\\ [y]=ab}}\frac{\xi(ab)}{\xi(a_0b_0)}\frac{\mathbb{L}(a_0b_0,\bar{x}_0ab)}{\#\{j_{\kappa_1,\kappa_2}\mid \kappa_1\neq s,ab\}+1}\Biggr]\\
=\ & \mathbb{P}\bigl[\mathbf{Z}_1=(s^{(1)},t^{(1)}),\dots,\mathbf{Z}_n=(s^{(n)},t^{(n)}),[\mathbf{x}_{n+1}]=a_0b_0\bigr]\frac{\xi(ab)}{\xi(a_0b_0)}\mathbb{L}(a_0b_0,\bar{x}_0ab)\\
=\ & \mathbb{P}\bigl[\mathbf{Y}_1=(s^{(1)},t^{(1)}),\dots,\mathbf{Y}_n=(s^{(n)},t^{(n)}),[\mathbf{W}_{n+1}]=a_0b_0\bigr]\frac{\xi(ab)}{\xi(a_0b_0)}\mathbb{L}(a_0b_0,\bar{x}_0ab)\\
=\ & \mathbb{P}\bigl[\mathbf{Y}_1=(s^{(1)},t^{(1)}),\dots,\mathbf{Y}_{n+1}=(s^{(n+1)},j_1),[\mathbf{W}_{n+1}]=a_0b_0,[\mathbf{W}_{n+2}]=ab\bigr].
\end{aligned}$$

Hence,

$$\begin{aligned}
&\mathbb{P}\bigl[\mathbf{Z}_1=(s^{(1)},t^{(1)}),\dots,\mathbf{Z}_{n+1}=(s^{(n+1)},t^{(n+1)}),[\mathbf{x}_{n+2}]=ab\bigr]\\
=\ & \sum_{a_0b_0\in\mathcal{A}^2}\mathbb{P}\bigl[\mathbf{Z}_1=(s^{(1)},t^{(1)}),\dots,\mathbf{Z}_{n+1}=(s^{(n+1)},t^{(n+1)}),[\mathbf{x}_{n+1}]=a_0b_0,[\mathbf{x}_{n+2}]=ab\bigr]\\
=\ & \sum_{a_0b_0\in\mathcal{A}^2}\mathbb{P}\bigl[\mathbf{Y}_1=(s^{(1)},t^{(1)}),\dots,\mathbf{Y}_{n+1}=(s^{(n+1)},t^{(n+1)}),[\mathbf{W}_{n+1}]=a_0b_0,[\mathbf{W}_{n+2}]=ab\bigr]\\
=\ & \mathbb{P}\bigl[\mathbf{Y}_1=(s^{(1)},t^{(1)}),\dots,\mathbf{Y}_{n+1}=(s^{(n+1)},t^{(n+1)}),[\mathbf{W}_{n+2}]=ab\bigr].
\end{aligned}$$

This proves Equation (C.2) for all $n\in\mathbb{N}$, all $ab\in\mathcal{A}^2$ and all $(s^{(1)},t^{(1)}),\dots,(s^{(n)},t^{(n)})\in\mathcal{W}_\pi$. Finally, we obtain:

$$\begin{aligned}
&\mathbb{P}\bigl[\mathbf{Z}_1=(s^{(1)},t^{(1)}),\dots,\mathbf{Z}_{n+1}=(s^{(n+1)},t^{(n+1)})\bigr]\\
=\ & \sum_{ab\in\mathcal{A}^2}\mathbb{P}\bigl[\mathbf{Z}_1=(s^{(1)},t^{(1)}),\dots,\mathbf{Z}_{n+1}=(s^{(n+1)},t^{(n+1)}),[\mathbf{x}_{n+2}]=ab\bigr]\\
=\ & \sum_{ab\in\mathcal{A}^2}\mathbb{P}\bigl[\mathbf{Y}_1=(s^{(1)},t^{(1)}),\dots,\mathbf{Y}_{n+1}=(s^{(n+1)},t^{(n+1)}),[\mathbf{W}_{n+2}]=ab\bigr]\\
=\ & \mathbb{P}\bigl[\mathbf{Y}_1=(s^{(1)},t^{(1)}),\dots,\mathbf{Y}_{n+1}=(s^{(n+1)},t^{(n+1)})\bigr].
\end{aligned}$$

This finishes the proof. □

References

[1] A. Avez, *Entropie des groupes de type fini*, C.R.Acad.Sci.Paris Sér. A-B **275** (1972), A1363–A1366. MR-0324741

[2] I. Benjamini and Y. Peres, *Tree-indexed random walks on groups and first passage percolation*, Probab. Theory Related Fields **98** (1994), no. 1, 91–112. MR-1254826

[3] S. Blachère, P. Haïssinsky, and P. Mathieu, *Asymptotic entropy and Green speed for random walks on countable groups*, Ann. Probab. **36** (2008), no. 3, 1134–1152. MR-2408585

[4] Tullio Ceccherini-Silberstein and Wolfgang Woess, *Context-free pairs of groups I: Context-free pairs and graphs*, European J. Combin. **33** (2012), 1449–1466. MR-2923462

[5] T. Cover and J. Thomas, *Elements of information theory*, 2nd ed., Wiley & Sons, 2006. MR-2239987

[6] Lorenz Gilch and Francois Ledrappier, *Regularity of the drift and entropy of random walks on groups*, Publ. Mat. Urug. **14** (2013), 147–158. MR-3235351

[7] Lorenz A. Gilch, *Rate of escape of random walks on regular languages and free products by amalgamation of finite groups*, Discrete Math. Theor. Comput. Sci. Proc. (2008), 405–420. MR-2508803

[8] Lorenz A. Gilch, *Asymptotic entropy of random walks on free products*, Electron. J. Probab. **16** (2011), 76–105. MR-2749773

[9] Lorenz A. Gilch and Sebastian Müller, *Random walks on directed covers of graphs*, J. Theoret. Probab. **24** (2009), no. 1, 118–149. MR-2782713

[10] S. Gouëzel, *Martin boundary of measures with infinite support in hyperbolic groups*, Ann. Probab (to appear). MR-3395464

[11] P. Haïssinsky, P. Mathieu, and S. Müller, *Renewal theory for random walks on cocompact Fuchsian groups*, Preprint (2012).

[12] Guangyue Han and Brian Marcus, *Analyticity of entropy rate of hidden Markov chains*, IEEE Trans. Inform. Theory **52** (2006), no. 12, 5251–5266. MR-2300690

[13] J. Hopcraft and J. Ullman, *Introduction to automata theory, languages, and computation*, Addison-Wesley, Reading MA, 1979. MR-0645539

[14] Vadim Kaimanovich and Wolfgang Woess, *Boundary and entropy of space homogeneous Markov chains*, Ann. Probab. **30** (2002), 323–363. MR-1894110

[15] J.F.C. Kingman, *The ergodic theory of subadditive processes*, J. Roy. Statist. Soc., Ser. B **30** (1968), 499–510. MR-0254907

[16] Steven Lalley, *Random walks on regular languages and algebraic systems of generating functions*, Algebraic Methods in Statistics and Probability, Contemp. Math. **287** (2000), 201–230. MR-1873677

[17] F. Ledrappier, *Analyticity of the entropy for some random walks*, Groups Geom. Dyn. **6** (2012), 317–333. MR-2914862

[18] F. Ledrappier, *Regularity of the entropy for random walks on hyperbolic groups*, Ann. Probab. **41** (2013), no. 5, 3582–3605. MR-3127892

[19] V. Malyshev, *Stabilization laws in the evolution of a random string*, Problems Inform. Transmission **30** (1995), 260–274. MR-1299692

[20] Pierre Mathieu, *Differentiating the entropy of random walks on hyperbolic groups*, Ann. Probab. **43** (2015), no. 1, 166–187. MR-3298471

[21] D.E. Muller and P.E. Schupp, *Groups, the theory of ends and context-free languages*, J. Comput. System Sci. **26** (1983), 295–310. MR-0710250

[22] D.E. Muller and P.E. Schupp, *The theory of ends, pushdown automata, and second-order logic*, Theoret. Comput. Sci. **37** (1985), 51–75. MR-0796313

[23] W. Woess, *Context-free pairs of groups. II – Cuts, tree sets, and random walks*, Discr. Math. **312** (2012), 157–173. MR-2852518